Juegos Matemáticos

Investigación y Ciencia (2001-2008)

Juan M.R. Parrondo

Juegos Matemáticos (2001-2008)

1. Perder+perder=ganar. Juegos paradójicos (Julio 2001).
2. Teoría de la información y juegos de preguntas sí/no (Agosto 2001).
3. Experimentos con compresores de ficheros (Septiembre 2001).
4. Las matemáticas de la opinión pública (Octubre 2001).
5. Las matemáticas del aprendizaje y la generalización (Noviembre 2001).
6. Juegos cuánticos (Diciembre 2001).
7. Información y juegos de azar: el problema de Monty Hall y la paradoja de los dos sobres (Enero 2002).
8. La paradoja de los dos sobres (Febrero 2002).
9. Paradojas democráticas (Marzo 2002).
10. Ventajas engañosas (Abril 2002).
11. Jugar con opciones y futuros (Mayo 2002).
12. Juegos equitativos con dados y monedas trucadas (Junio 2002).
13. Caos, determinismo y voluntad (Julio 2002).
14. Repartir escasez (Agosto 2002).
15. Paradojas y atascos de tráfico (Septiembre 2002).
16. Monedas, balanzas e información (Octubre 2002).
17. Ruletas, monedas y entropía (Noviembre 2002).
18. La misteriosa ley del primer dígito (Diciembre 2002).
19. El número de oro (Enero 2003).
20. Números y palabras (Febrero 2003).
21. Más sobre números y palabras (Marzo 2003).
22. Fluctuaciones fatales (Abril 2003).
23. El examen inesperado y la teoría de juegos (Mayo 2003).
24. Ruidos reveladores (Junio 2003).
25. La paradoja de Simpson (Julio 2003).
26. Un mundo sin números (Agosto 2003).
27. El problema del secador de manos (Septiembre 2003).
28. La paradoja de la Biblioteca de Babel (Octubre 2003).
29. Zenón y los camellos (Noviembre 2003).
30. Cita a ciegas (Diciembre 2003).
31. La frecuencia fantasma (Enero 2004).
32. Las ventajas de la solidaridad (Febrero 2004).
33. La teoría matemática de la consonancia (Marzo 2004).
34. Cuestión de escala (Abril 2004).
35. Matemáticas electorales (Mayo 2004).
36. La paradoja del autostopista (Junio 2004).
37. Matemáticas sostenibles (Julio 2004).
38. El reparto de poder en la Unión Europea (Agosto 2004).
39. Democracia ineficiente (Septiembre 2004).
40. Más sobre el reparto de poder (Octubre 2004).
41. Numerogooglia (Noviembre 2004).
42. Calculistas prodigiosos (Diciembre 2004).
43. El número mayor y la información misteriosa (Enero 2005).
44. Problemas de aparcamiento (Febrero 2005).

45. La dote del sultán (Marzo 2005).
46. Fósiles y lotería (Abril 2005).
47. Sorteos polémicos (Mayo 2005).
48. La forma de un iceberg (Junio 2005).
49. ¿Amigos para siempre? (Julio 2005).
50. Cribas y números primos (Agosto 2005).
51. Más paradojas de alternancia (Septiembre 2005).
52. Hagan sus apuestas (Octubre 2005).
53. Quién ríe el último... (Noviembre 2005).
54. Finalmente... sudoku (Diciembre 2005).
55. ¿Hay quien dé más? (enero 2006).
56. Incentivar la sinceridad (Febrero 2006).
57. El número h (Marzo 2006).
58. Caos, recurrencia y consonancia musical (Abril 2006).
59. El espacio-tiempo (Mayo 2006).
60. Espacio-tiempo y azar (Junio 2006).
61. Otras formas de contar (Julio 2006).
62. Ganancia segura (Agosto 2006).
63. Medir la desigualdad (Septiembre 2006).
64. El juego del ultimátum (Octubre 2006).
65. Los dados misteriosos y la razón áurea (Noviembre 2006).
66. La joya oculta (Diciembre 2006).
67. Los logaritmos de Brigss (Enero 2007).
68. La paradoja de San Petersburgo (Febrero 2007).
69. La paradoja de San Petersburgo y la teoría de la utilidad (Marzo 2007).
70. Loterías y decisiones (Abril 2007).
71. La asombrosa fórmula de Tupper (Mayo 2007).
72. Pensamiento formal y pensamiento concreto (Junio 2007).
73. Sutilezas estadísticas (Julio 2007).
74. Móviles y vectores (Agosto 2007).
75. Números pseudoaleatorios (Septiembre 2007).
76. Más sobre números aleatorios (Octubre 2007).
77. Carreras cuadriculadas (Noviembre 2007).
78. El caso de la moneda cambiada (Diciembre 2007).
79. El juego de las avalanchas (Enero 2008).
80. Sorpresas termodinámicas (Febrero 2008).
81. Estimaciones (Marzo 2008).
82. Cifras y letras (Abril 2008).
83. Encuestas electorales (Mayo 2008).
84. El problema de los tres dioses (Junio 2008).
85. Piensa un número (Julio 2008).

Temas

Probabilidad

1. Perder+perder=ganar. Juegos paradójicos (Julio 2001).
7. Información y juegos de azar: el problema de Monty Hall y la paradoja de los dos sobres (Enero 2002).
8. La paradoja de los dos sobres (Febrero 2002).
10. Ventajas engañosas (Abril 2002).
22. Fluctuaciones fatales (Abril 2003).
24. Ruidos reveladores (Junio 2003).
31. La frecuencia fantasma (Enero 2004).
36. La paradoja del autostopista (Junio 2004).
43. El número mayor y la información misteriosa (Enero 2005).
44. Problemas de aparcamiento (Febrero 2005).
45. La dote del sultán (Marzo 2005).
50. Cribas y números primos (Agosto 2005).
51. Más paradojas de alternancia (Septiembre 2005).
52. Hagan sus apuestas (Octubre 2005).
53. Quién ríe el último... (Noviembre 2005).
65. Los dados misteriosos y la razón áurea (Noviembre 2006).
68. La paradoja de San Petersburgo (Febrero 2007).
69. La paradoja de San Petersburgo y la teoría de la utilidad (Marzo 2007).
73. Más sobre números aleatorios (Octubre 2007).
75. El caso de la moneda cambiada (Diciembre 2007).

Estadística

18. La misteriosa ley del primer dígito (Diciembre 2002).
20. Números y palabras (Febrero 2003).
21. Más sobre números y palabras (Marzo 2003).
25. La paradoja de Simpson (Julio 2003).
41. Numerogooglia (Noviembre 2004).
46. Fósiles y lotería (Abril 2005).
47. Sorteos polémicos (Mayo 2005).
61. Otras formas de contar (Julio 2006).
70. Sutilezas estadísticas (Julio 2007).
80. Encuestas electorales (Mayo 2008).

Teoría de la información

2. Teoría de la información y juegos de preguntas sí/no (Agosto 2001).
3. Experimentos con compresores de ficheros (Septiembre 2001).
5. Las matemáticas del aprendizaje y la generalización (Noviembre 2001).
12. Juegos equitativos con dados y monedas trucadas (Junio 2002).
16. Monedas, balanzas e información (Octubre 2002).
28. La paradoja de la Biblioteca de Babel (Octubre 2003).
72. Números pseudoaleatorios (Septiembre 2007).

Teoría de juegos y economía y sociología

4. Las matemáticas de la opinión pública (Octubre 2001).
9. Paradojas democráticas (Marzo 2002).
14. Repartir escasez (Agosto 2002).
15. Paradojas y atascos de tráfico (Septiembre 2002).
11. Jugar con opciones y futuros (Mayo 2002).
23. El examen inesperado y la teoría de juegos (Mayo 2003).
30. Cita a ciegas (Diciembre 2003).
32. Las ventajas de la solidaridad (Febrero 2004).
35. Matemáticas electorales (Mayo 2004).
38. El reparto de poder en la Unión Europea (Agosto 2004).
39. Democracia ineficiente (Septiembre 2004).
40. Más sobre el reparto de poder (Octubre 2004).
49. ¿Amigos para siempre? (Julio 2005).
55. ¿Hay quien dé más? (enero 2006).
56. Incentivar la sinceridad (Febrero 2006).
57. El número h (Marzo 2006).
63. Medir la desigualdad (Septiembre 2006).
64. El juego del ultimátum (Octubre 2006).
67. Loterías y decisiones (Abril 2007).

Física

6. Juegos cuánticos (Diciembre 2001).
17. Ruletas, monedas y entropía (Noviembre 2002).
48. La forma de un iceberg (Junio 2005).
59. El espacio-tiempo (Mayo 2006).
60. Espacio-tiempo y azar (Junio 2006).
76. El juego de las avalanchas (Enero 2008).
77. Sorpresas termodinámicas (Febrero 2008).
78. Estimaciones (Marzo 2008).

Filosofía e historia

13. Caos, determinismo y voluntad (Julio 2002).
19. El número de oro (Enero 2003).
26. Un mundo sin números (Agosto 2003).
66. La joya oculta (Diciembre 2006).
67. Los logaritmos de Brigss (Enero 2007).
69. Pensamiento formal y pensamiento concreto (Junio 2007).

Análisis matemático, álgebra y optimización

27. El problema del secador de manos (Septiembre 2003).
29. Zenón y los camellos (Noviembre 2003).
37. Matemáticas sostenibles (Julio 2004).
71. Móviles y vectores (Agosto 2007).

Música

33. La teoría matemática de la consonancia (Marzo 2004).
34. Cuestión de escala (Abril 2004).
58. Caos, recurrencia y consonancia musical (Abril 2006).

Matemática recreativa

JUEGOS MATEMÁTICOS

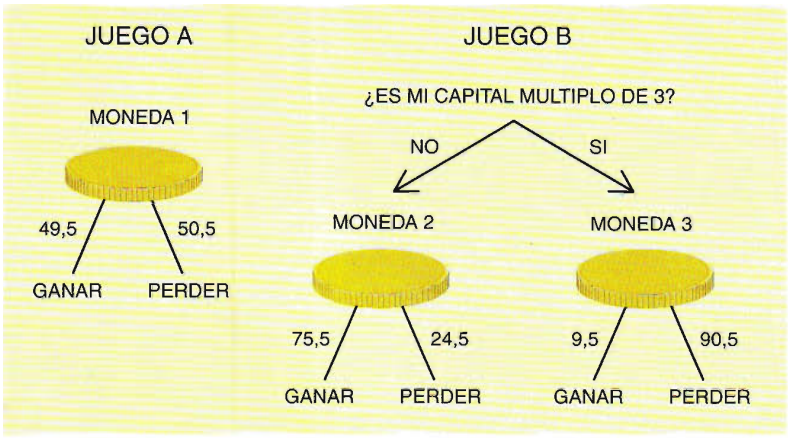
Perder + perder = ganar. Juegos de azar paradójicos

El estudio de ciertas propiedades del movimiento browniano —el movimiento azaroso que experimentan pequeñas partículas debido a las colisiones con las moléculas del fluido en que están inmersas y cuya teoría fue establecida por Einstein en 1905— ha inspirado recientemente una curiosa paradoja. Se trata de dos juegos de azar muy simples diseñados de modo tal que el jugador, en promedio, pierde en ambos. La paradoja consiste en que esta tendencia se invierte cuando los juegos se alternan, en cuyo caso el jugador tiene una tendencia ganadora constante. Veamos en detalle las reglas de estos dos juegos.

El primero de ellos —lo llamaremos *juego A*— es similar a apostar una cierta cantidad, digamos 1 euro, a rojo o negro en la ruleta de un casino: ganamos 1 con una probabilidad ligeramente inferior al 50 % y perdemos 1 con una probabilidad ligeramente superior al 50 %, ya que con el cero gana siempre la banca. Supongamos que se gana con una probabilidad del 49,5 % y se pierde con una probabilidad del 55,5 % (estas probabilidades no coinciden exactamente con las de la ruleta, pero nos permitirán describir de modo sencillo la paradoja).

En el segundo juego —lo llamaremos *juego B*—, en cada turno también ganamos o perdemos 1 , pero ahora las probabilidades dependen de lo que llevamos ganado hasta el momento (que puede ser una cantidad negativa): si lo que llevamos ganado —lo llamaremos *el capital*— es múltiplo de 3, entonces ganamos 1 con probabilidad 9,5 %; si el capital no es múltiplo de 3, la probabilidad de ganar es del 74,5 %. Las reglas de los dos juegos se muestran en la figura 1.

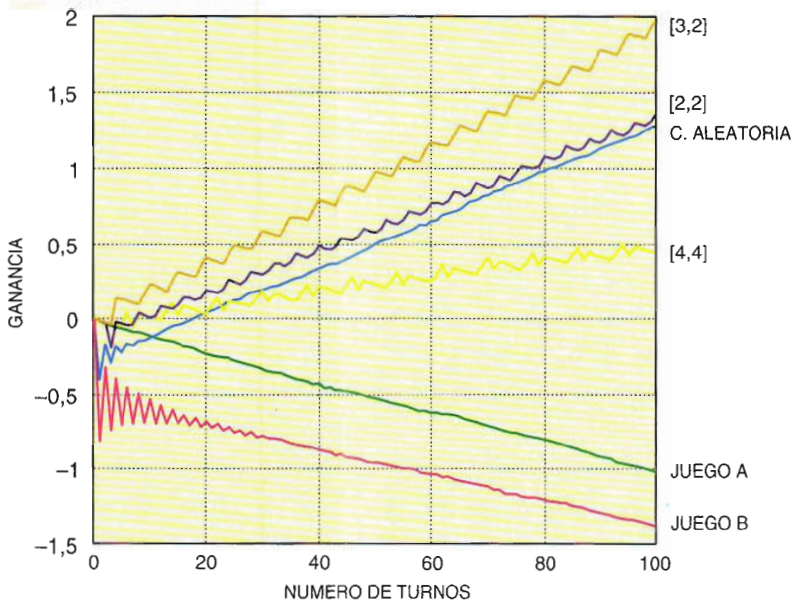
Ambos juegos son desfavorables, es decir, la tendencia promedio es perdedora (esto es evidente en el caso del juego *A*; el juego *B* requiere un análisis más detallado, pero puede también demostrarse que es desfavorable en promedio). Sin embargo, ciertas combinaciones de los dos juegos tienen una tendencia media ganadora. Este comportamiento inesperado, que se conoce como *Paradoja de Parrondo*, se observa en la figura 2, en la que se muestra el capital promedio de 5000 jugadores independientes en función del número de turnos jugados. Las curvas descendentes indican los casos en que se juega a *A* o *B* separadamente. Las curvas ascendentes corresponden a distintas combinaciones: [2,2] significa que jugamos la sucesión *AABBAABB*..., [3,2] la sucesión *AAABBAAABB*..., etc. Finalmente, la "combinación aleatoria" consiste en que, en cada turno, elegimos completamente al azar a cuál de los dos juegos jugamos.



1. Reglas de los juegos paradójicos

No es difícil entender por qué ocurre este comportamiento paradójico. Obsérvese que, en el juego *B*, la probabilidad de ganar es muy baja cuando el capital es múltiplo de 3 y bastante alta cuando no lo es. Podríamos decir que el juego *B* contiene dos tendencias, una favorable y otra desfavorable. Las probabilidades están elegidas de modo tal, que el juego es ligeramente perdedor, es decir, la tendencia desfavorable domina sobre la favorable. Lo que ocurre cuando combinamos ambos juegos es que el *A* invierte este dominio. A pesar de que el juego *A* es en sí ligeramente desfavorable, hace que el capital sea menos veces múltiplo de 3 y que, por tanto, juguemos más veces con la probabilidad alta de ganar, la del 74,5 %. El juego *A*, en la alternancia, cumple un doble papel: añade por un lado su propia tendencia desfavorable, pero, por otro lado, refuerza la tendencia favorable del juego *B*. El resultado global de este doble papel es, en casos como los de la figura, invertir la tendencia y dar lugar al comportamiento paradójico.

Existe otro mecanismo interesante tras la paradoja. Lo podemos ilustrar con el siguiente ejemplo en la



2. Ganancia media en cada uno de los juegos y en sus combinaciones

secuencia *AABBAABB*... Supongamos que jugamos *A* con capital múltiplo de tres, 15 por ejemplo, y que perdemos los dos turnos de *A*. En tal caso, retomamos el juego *B* con un capital igual a 13 , para el que la probabilidad de ganar es alta. Por lo tanto, lo más probable es que el capital vuelva hacia 15 en los dos turnos de *B*. Podemos decir que el juego *B* no permite que las pérdidas de *A* se "consoliden", pero sí permite que lo hagan las ganancias. De modo que, aunque *A* sea desfavorable, y sea mayor la frecuencia de turnos perdidos que la de turnos ganados, el juego *B* se encarga de "destruir" las fluctuaciones negativas y de "consolidar" las fluctuaciones positivas. Decimos entonces que el juego *B* actúa como un *rectificador* o una *ratchet*, término inglés que denota una rueda de dientes asimétricos que sólo puede girar en un sentido y que se utiliza, entre otras muchas cosas, para *rectificar* el movimiento fluctuante de la muñeca y dar cuerda a un reloj de pulsera. Este efecto rectificador de fluctuaciones genera un movimiento sistemático en una dirección sin necesidad de ninguna fuerza, simplemente "esperando" a que ocurran las fluctuaciones adecuadas.

Los dos mecanismos que hemos descrito: la reorganización de tendencias y la rectificación de fluctuaciones, son bastante simples y básicos, de modo que pueden estar presentes en muchas situaciones. La paradoja nos muestra que alternar dinámicas aleatorias puede dar lugar a efectos inesperados, contrarios a los de cada dinámica por separado, y esta afirmación general sugiere nuevas líneas de investigación en física, química, biología o economía, en las que se estudie el comportamiento de sistemas diversos sometidos a una alternancia de temperaturas, campos eléctricos u otras condiciones externas.

Desde la publicación de la paradoja se han descrito algunas modificaciones, como juegos en donde las reglas no dependen del capital y juegos colectivos. Todas estas investigaciones indican que la paradoja no se limita a situaciones particulares. Recientemente, Paul Davies ha apuntado que podría tener alguna relevancia en la explicación del origen de la vida, ya que ésta requiere una sutil combinación de azar y necesidad y los juegos nos muestran mecanismos sencillos en donde el azar es transformado en un movimiento o tendencia sistemática, es decir, en un cierto orden.

Juan MR Parrondo
Facultad de Físicas, Universidad Complutense, Madrid

JUEGOS MATEMÁTICOS

Juan MR Parrondo

¿Cuál es la mejor estrategia en un juego de preguntas sí/no?

En una capital de provincias se ha descubierto un virus letal que se propaga por vía sanguínea. El Instituto Nacional de la Salud establece un protocolo de prueba, que detecta la presencia del virus. La prudencia aconseja que toda donación de sangre se someta a la prueba.

Pero los reactivos que se utilizan son muy caros. Además, la incidencia del virus en la población es aún baja, de un exiguo 0,01 %. A los responsables de los bancos de sangre no se les escapa un viejo truco: mezclar la sangre de un grupo de donantes y pasar la mezcla por la prueba. Si sale negativo, ausencia de virus, podemos asegurar que ninguno de los integrantes de la mezcla está infectado. Si sale positivo, presencia de virus, se procede a un análisis individualizado de cuantos componen el grupo de la mezcla. Pero, ¿cuándo podemos decir que el número de integrantes del grupo de donantes aunado para la mezcla es el óptimo? Dicho de otro modo: ¿cuál es el tamaño de la mezcla exigido para realizar, en media, el menor número de pruebas? (El problema así planteado puede resolverse con determinados recursos probabilísticos, que hoy dejaremos de lado.)

Con toda lógica los responsables de los bancos piensan que el proceso de mezclas puede aplicarse de nuevo cuando la prueba de la primera mezcla es positiva. En este caso tenemos un cierto número de donantes, entre los cuales sabemos que existe al menos uno infectado. Y podemos someter a otra prueba a ciertas mezclas de estos donantes. ¿Cuál será entonces el diseño óptimo de las nuevas mezclas? El problema ahora es bastante más complicado. Sin embargo, la solución se revela de una extraordinaria sencillez. Guarda relación con la teoría de la información y con dos aplicaciones suyas, la codificación de mensajes y la compresión de ficheros de datos.

El problema de la mezcla de sangre es similar a los juegos en los que se tiene que adivinar algo —un personaje, un objeto de la habitación— mediante preguntas que sólo admiten una respuesta, afirmativa o negativa. ¿Cuál es la mejor estrategia en este tipo de juegos? Sin advertirlo, la mayoría de la gente encuentra la óptima de modo un intuitivo, al menos en las primeras fases del juego. Sea por caso adivinar un personaje. ¿Es hombre o mujer? ¿español o extranjero?, etcétera. Preguntas para las cuales la probabilidad de la respuesta afirmativa es igual a la de la respuesta negativa, es decir, 1/2.

De acuerdo con la teoría de la información la pregunta óptima es aquella cuya respuesta tiene una probabilidad 1/2 de ser SÍ, y una probabilidad 1/2 de ser NO. Receta sencilla que vale para el juego de adivinación y para el problema de las mezclas de sangre.

Imaginemos, por ejemplo, que nos pidieran adivinar un número entre el 1 y el 16. Podríamos comenzar preguntando si es par o impar. De este modo descartamos 8 números y nos quedamos con 8. La siguiente pregunta tendría que descartar 4 números. En efecto, si nos han respondido que el número es par, podríamos ahora inquirir: ¿es múltiplo de cuatro? La tercera pregunta debe descartar 2 números; por último, la cuarta decidirá entre los dos que quedan. Con esta estrategia siempre acertamos el número tras cuatro preguntas, no importa cuál sea la cifra.

Podemos diseñar estrategias mejores para ciertos números. Si empezamos preguntando ¿es el 16?, la estrategia sería muy buena en el caso de que el número pensado fuera el 16, pero habríamos prácticamente desperdiciado el turno si no lo fuera; tras esa pregunta directa, sólo se ha descartado un número de los 16 posibles y la incertidumbre inicial apenas ha disminuido.

No resulta difícil demostrar que la estrategia primera —la de adivinar un número tras cuatro preguntas, con una probabilidad 1/2 de respuesta afirmativa en cada una— es, en media, la que requiere menor número de preguntas. Esta estrategia óptima se utiliza en un

1. Cartones para adivinar un número del 1 al 16

La teoría de la información

¿Hay alguna forma de cuantificar la incertidumbre, lo que desconocemos de una determinada situación? La teoría de la información de Claude Shannon nos ofrece una respuesta. Si en una situación, un experimento, un juego de adivinación, etcétera, se pueden dar diferentes instancias, 1,2,3,..., con probabilidades p_1, p_2, p_3,..., entonces la incertidumbre de esa situación es:

$$H = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - p_3 \log_2 p_3 - \ldots$$

La unidad que mide la incertidumbre, tal y como está definida en esta fórmula, es el bit. En el ejemplo de la adivinación de un número del 1 al 16, la incertidumbre inicial es H = 4 bits, ya que todos los números son igualmente probables (siempre que la persona que elige el número no tenga algún sesgo, predilección por los números pares o aversión por el 13, por ejemplo, y que nosotros conozcamos dicho sesgo) y su probabilidad es 1/16. Como el logaritmo en base 2 de 1/16 es –4, el resultado de la suma es 4.

Después de la primera pregunta, el lector puede comprobar que la incertidumbre es de 3 bits; después de la segunda pregunta, de 2 bits. La tercera reduce la incertidumbre hasta 1 bit y, finalmente, la última pregunta elimina toda la incertidumbre y hace que H sea igual a cero.

Como cabe esperar de nuestra noción intuitiva de incertidumbre, cada pregunta la reduce. Pero la definición de Shannon es cuantitativa; podemos calcular cuánto se reduce la incertidumbre. Por ejemplo, después de la pregunta ¿es 16 el número que has pensado?, si respondemos afirmativamente la incertidumbre se hace cero, mientras que con una respuesta negativa la incertidumbre es 3,9 bits. Esta pregunta apenas reduce la incertidumbre inicial si el número secreto no es 16. Para cuantificar el riesgo que estamos asumiendo con esta pregunta, lo mejor es calcular la incertidumbre media después de realizarla. Con una probabilidad 1/16, el número secreto es 16 y la incertidumbre se habrá hecho cero. Con una probabilidad 15/16 el número secreto no es 16 y la incertidumbre se habrá reducido a 3,9 bits. La incertidumbre media después de la pregunta es, por tanto: 0 por 1/16 más 3,9 por 15/16, es decir, 3,66 bits. De acuerdo con la teoría de Shannon resulta más eficaz preguntar si es par el número pensado que correr el riesgo.

Cada pregunta reduce la incertidumbre que rodea la situación, reducción que depende de la pregunta realizada. Pues bien, llamamos información media suministrada por la pregunta a los bits en que ha sido capaz de reducir la incertidumbre:

$$H_{\text{media,después}} = H_{\text{antes}} - I_{\text{pregunta}}$$

o, dicho en lenguaje más llano, la incertidumbre que tenemos después de conocer la respuesta a nuestra pregunta es la incertidumbre que teníamos antes menos la información suministrada por la pregunta. Si hay un ejemplo fascinante de cómo hacer una teoría cuantitativa a partir de nociones cualitativas y poco precisas, éste es la teoría de la información de Shannon.

Pero se puede ir aún más allá y demostrar que una pregunta cuya probabilidad de ser respondida afirmativamente sea p y de ser respondida negativamente sea 1 – p, suministra una información media:

$$I_{\text{pregunta}} = -p \log_2 p - (1 - p) \log_2 (1 - p)$$

Una fórmula que se parece mucho a la de la incertidumbre, si bien merece una interpretación distinta (y confundirla es algo que, lamentablemente, ocurre incluso en algunos textos de teoría de la información). El lector familiarizado con probabilidades y logaritmos puede demostrar este resultado. Déjenme dar sólo unas indicaciones para quienes se atrevan con la demostración: imaginen que la pregunta es ¿pertenece el número que has pensado a un conjunto A? Entonces p tiene que ser necesariamente la suma de las probabilidades de los elementos de A, y 1 – p la suma de las probabilidades de los elementos que no pertenecen a A. Por otra parte, si la respuesta es afirmativa, las nuevas probabilidades son pi/p y, si la respuesta es negativa, las nuevas probabilidades serán pi/(1 – p).

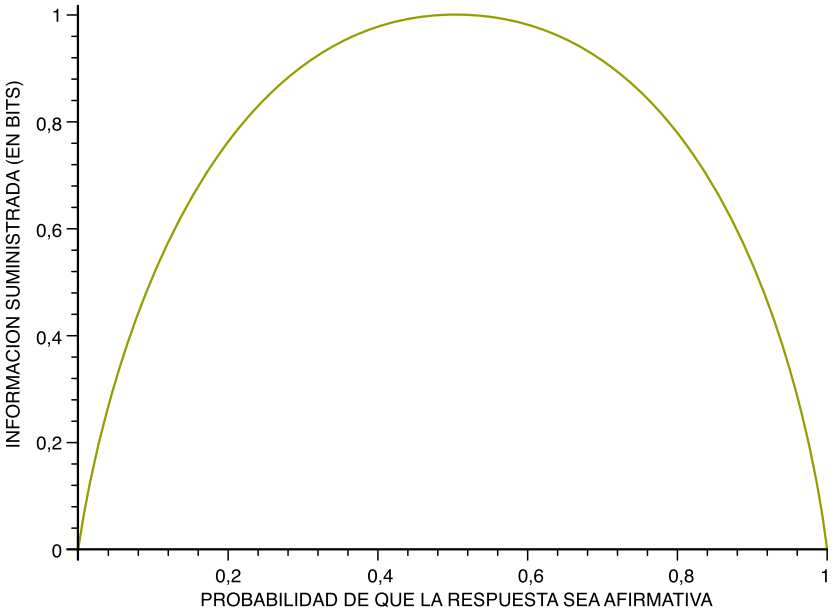
La fórmula de la información media suministrada por la pregunta tiene algo de mágico. Sólo depende de p, la probabilidad de que la pregunta sea contestada afirmativamente. En la figura 2 se dibuja la forma de esta curva. Toma su valor máximo, 1 bit, cuando p = 1/2. Las preguntas que aportan la información máxima son las que tienen la misma probabilidad de ser contestadas afirmativa y negativamente. La información ofrecida es 1 bit. Por eso, en el juego de adivinación del número, la incertidumbre inicial de 4 bits se reduce en cada pregunta en una unidad y podemos conocer el número después de la cuarta pregunta. Ratifica la gráfica que, para p = 1/16, la información media suministrada es 0,34 bits, como hemos deducido un poco más arriba.

Tenemos finalmente un resultado muy interesante: si estamos ante una situación cuya incertidumbre es de H bits, y si podemos realizar una serie preguntas sí/no con probabilidad 1/2 de ser respondidas afirmativamente, es decir, preguntas que suministran 1 bit de información, entonces el número medio de preguntas necesarias para eliminar la incertidumbre será H. La teoría de la información nos dice, por tanto, cómo debemos hacer las preguntas y cuántas se requieren.

juego de adivinación bastante difundido. Participan dos jugadores, llamémosles Benito y Alicia. Benito piensa en un número entre 1 y 16, que Alicia debe adivinar. Le muestra cuatro cartones, como los de la figura 1. Alicia debe averiguar en cuál de ellos se encuentra el número pensado.

Con esta información no le costará mucho. Bastará con que Alicia sume 8 si Benito le dice que está en el primer cartón, más 4 si está en el segundo, más 2 si se halla en el tercero, más 1 si está en el último cartón. Finalmente, a esta suma se le añade 1 y se obtiene el número secreto. Por ejemplo, si el número es el 11, Benito le dirá que se encuentra en el primero y tercer cartones. Sumará entonces 8 + 2 + 1 y obtendrá el número deseado.

El lector habrá adivinado enseguida que los cartones no son más que la representación binaria del número secreto menos 1. En el



2. Información media suministrada por una pregunta en función de la probabilidad de ser respondida afirmativamente

caso del 11, la representación binaria de 10 es precisamente 1010. En el primer cartón se encuentran los números cuya representación binaria tiene un 1 en el cuarto dígito; en el segundo cartón los que tienen un 1 en el tercer dígito, etcétera.

Detrás de este truco tan sencillo se esconde una idea interesante: la tarea de adivinar algo a través de preguntas sí/no equivale a la de *codificar* ese algo en representación binaria. Si llevamos esta equivalencia un poco más allá, nos percataremos de que encontrar la estrategia óptima en el juego de preguntas, es decir, la estrategia que minimiza en media el número de preguntas, equivale a hallar la codificación más corta.

Precisamente la búsqueda de codificaciones cortas fue uno de las primeras aplicaciones de la teoría de la información de Claude Shannon. Por eso esta teoría tiene que ver con asuntos de apariencia tan dispar como los juegos de preguntas sí/no, el problema de la seguridad de las transfusiones de sangre o los algoritmos que comprimen los ficheros de su ordenador antes de enviarlos a través de Internet.

La teoría de la información nos proporciona una definición cuantitativa de la incertidumbre o el "desconocimiento" de una situación, y, por supuesto de la información que suministra la respuesta a una pregunta acerca de dicha situación. En el recuadro se ofrecen detalles técnicos de la teoría y una gráfica en donde se muestra la información media suministrada por una pregunta sí/no en función de la probabilidad de que se responda afirmativamente.

Como vemos en el gráfico, la información máxima la suministran las preguntas en las que la respuesta afirmativa y negativa tienen la misma probabilidad, es decir, 1/2. Por tanto, la estrategia óptima en un juego de adivinación consistirá en atenerse a este tipo de preguntas. Nos dice la teoría que el número mínimo de preguntas sí/no necesarias para adivinar es igual a la *incertidumbre* inicial y nos proporciona una fórmula para calcular esta incertidumbre.

Vayamos con otra aplicación, la compresión de ficheros. Un fichero es un conjunto de datos binarios. Si agrupamos los datos en bloques de 20 dígitos binarios, tendremos 220 posibles bloques. El fichero puede entonces considerarse como una lista de "palabras" extraídas de este diccionario que contiene un millón de palabras. Cada palabras es un conjunto de dígitos binarios.

Pero podemos construir otra representación binaria alternativa en la que las palabras más frecuentes estén representadas por cadenas de dígitos más cortas. De eso se ocupa un compresor de ficheros. Para ello aplica la teoría de la información. Igual que en el caso de los cartones, una representación binaria de cada una de estas palabras puede interpretarse como las respuestas a una serie de preguntas sí/no acerca de dicha palabra. Y la estrategia óptima de preguntas es equivalente a encontrar la representación que, en media, es más corta.

A la compresión de ficheros dedicaremos el artículo del mes próximo. En números sucesivos abordaremos también la solución del problema de las transfusiones de sangre en el caso más sencillo, en el que, tras una prueba positiva de la mezcla, se hace la prueba *individualmente* a cada integrante de la misma. La respuesta difiere de la dada por la teoría de la información, es decir, el tamaño óptimo de la mezcla no hace que la probabilidad de que la prueba resulte positiva sea 1/2. ¿Por qué? Porque la teoría de la información sólo puede aplicarse a la estrategia completa de preguntas o, en este caso, de mezclas. Es decir, sólo puede aplicarse cuando, después de la prueba positiva, analizamos nuevas "submezclas". En ese caso la mejor estrategia consiste en crear mezclas cuya probabilidad de contener el virus se cifre en 1/2. ¿Es esto todo lo que puede decir al respecto? Espero que los lectores me hagan llegar sus propias respuestas y sugerencias.

No se podría terminar un artículo sobre teoría de la información sin rendir un pequeño homenaje a su creador, Claude Shannon, que murió el pasado 24 de febrero a la edad de 84 años. Quienes deseen saber más sobre su vida, y también sobre su teoría, pueden visitar el excelente sitio web que los Laboratorios Bell, en los que Shannon desarrolló la mayor parte de su actividad, han dedicado a este tema: http://www.lucent.com/minds/infotheory.

parr-km0@zenon.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Experimentos con compresores de ficheros

Toda persona que utilice Internet se habrá dado cuenta de lo importante que es comprimir la información antes de enviarla a través de la red. Un nuevo modo de comprimir datos puede originar a veces a una auténtica revolución en este mundillo, como ha ocurrido con el formato *mp3*. Pero la compresión de datos es un problema matemático cuya importancia va más allá de estas aplicaciones y afecta a cuestiones tan fundamentales como la definición objetiva del azar.

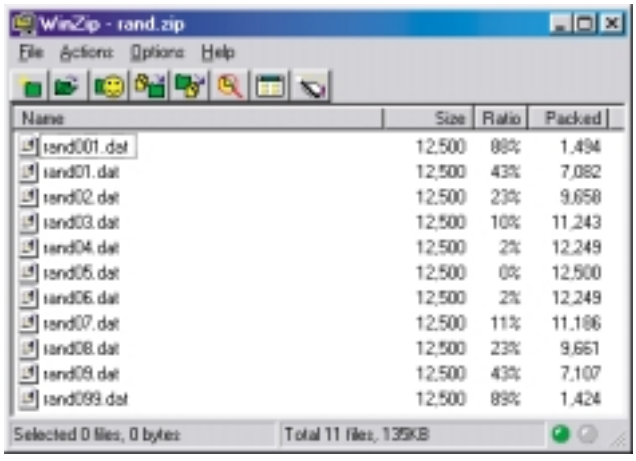
Claude Shannon demostró que el tamaño mínimo al que se puede reducir un fichero de datos es igual a su incertidumbre o entropía y mostró la manera de calcular esta entropía en casos sencillos [*véase* Juegos matemáticos del mes de agosto]. En este artículo mostraremos algunos experimentos sencillos que ilustran el teorema de Shannon y que nos ayudarán también a investigar cuánta información contenga una sonata de Beethoven o el *Ulises* de Joyce. Nos bastará para ello cualquier programa de compresión de ficheros para Windows, Macintosh o UNIX.

Uno de los algoritmos de compresión más conocido es el llamado de Lempel-Ziv (LZ). El programa *compress* del sistema operativo UNIX y todos los programas que generan ficheros *.arj*, *.zip* o *.gif* en Windows utilizan variantes suyas. Lo que hace el algoritmo es aprovechar de forma bastante simple las repeticiones de ciertas "frases" que aparecen en una cadena de bit, es decir, en una cadena de dígitos 0 o 1. Para ello se fragmenta primero la cadena de modo que no aparezca la misma frase dos veces. Esto se consigue colocando comas de forma sucesiva: se coloca una coma después del primer dígito; la siguiente se coloca de modo que el fragmento entre comas resultante sea el más corto posible que no haya aparecido antes. Por ejemplo, en la cadena de bit:

101110001010101010100

la fragmentación proporcionará:

1,0,11,10,00,101,01,010,100



1. El resultado de comprimir ficheros con cien mil bit aleatorios

El lector puede observar que cada fragmento es siempre la concatenación de un fragmento aparecido con anterioridad y de un bit 0 o 1. Los llamaremos fragmento prefijo y bit adicional, respectivamente. Por ejemplo, el tercer fragmento, 11, es el primero seguido de un 1; el sexto fragmento, 101, es el cuarto seguido de un 1; y así sucesivamente. Podemos ahora numerar los fragmentos y sustituir cada uno de ellos por el número del prefijo y por el bit adicional. El 0 indicará por convención la falta de prefijo o "prefijo vacío". En nuestro ejemplo:

(0,1),(0,0),(1,1),(1,0),(2,0),(4,1),(2,1),(7,0),(4,0)

Finalmente se representan en binario los números que indican el prefijo para obtener una nueva cadena de bit. Veamos cómo se realiza este paso en nuestro ejemplo. Como hemos utilizado siete prefijos, se necesitan tres bit para representar cada uno de ellos:

(000,1),(000,0),(001,1),(001,0),(010,0),(100,1),(010,1),
(111,0),(100,0)

Ahora se pueden quitar los paréntesis y las comas y obtener así la cadena de bit comprimida:

000100000011001001001001010111101000

La descompresión es muy sencilla, puesto que la representación binaria del número del prefijo tiene siempre el mismo número de bit, tres en nuestro ejemplo. Sabemos entonces que cada fragmento está codificado por cuatro bit: tres para el prefijo y uno para el bit adicional.

La cadena resultante de nuestro ejemplo es bastante más larga que la original, en contra del propósito del algoritmo. Esto se debe a la escasa eficacia que éste tiene en el caso de cadenas muy cortas. Cuando se aplica a una cadena muy larga, los fragmentos cada vez son más grandes, mientras que el número de bit necesario para describirlos crece de manera más lenta. Por ejemplo, supongamos que una cadena de cien mil bit se ha dividido en mil fragmentos con un tamaño medio de cien bit (los primeros serán más cortos y los últimos más largos, probablemente). Para describir el prefijo no se necesitan más que 10 bit, ya que $2^{10} = 1024$. Por tanto cada fragmento estará codificado por 11 bit (los 10 del prefijo más el bit adicional), mientras que su longitud original era de 100 bit. Es decir, el algoritmo ha logrado comprimir este fichero de datos a un 10 % de su tamaño original.

Está claro que cuanto más largos sean los fragmentos, más eficaz será el algoritmo. Esto ocurre si

la cadena tiene muchas repeticiones. El caso extremo es una cadena con todo ceros. En este caso la fragmentación es muy simple:

0,00,000,0000,00000,...

El lector con algunos conocimientos de matemáticas puede comprobar que una cadena con n ceros se dividiría en un número de fragmentos aproximadamente igual a la raíz cuadrada de $2n$. Para 1 megabyte, que son 8 millones de bit, tendríamos 4000 fragmentos y necesitaríamos sólo 12 bit para describir el prefijo. Por tanto, necesitaríamos 13 bit para describir cada fragmento o 52.000 bit para describir el fichero entero. Como un byte son 8 bit, el fichero habría quedado reducido a 6,5 kilobyte.

El caso opuesto es una cadena de bit cuya fragmentación contenga todas las subcadenas posibles. Por ejemplo, hay 28 subcadenas de 8 bit. Si la fragmentación resultara en una concatenación de todas ellas, se necesitarían 8 bit para describir el prefijo y 9 para describir el fragmento, mientras que la longitud original de los fragmentos es de 8 bit. El algoritmo resultaría perfectamente inútil en este caso.

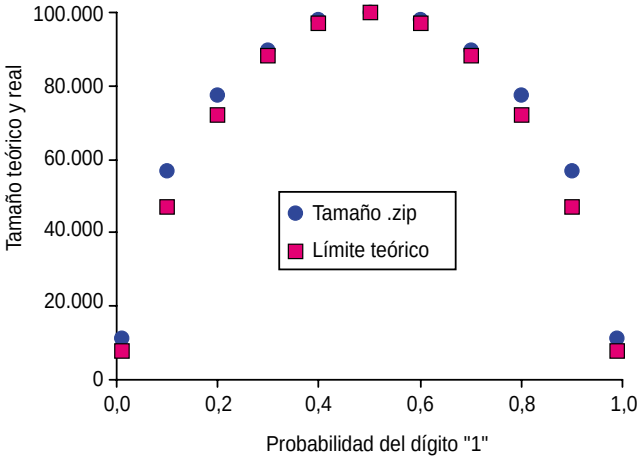
Como vemos, el algoritmo LZ es más eficaz cuanto más regular sea la cadena, cuantas más repeticiones contenga. Esta característica está de acuerdo con el teorema de Shannon: cuanto menor sea la entropía de una cadena, menos aleatoria y más regular es, por lo que puede comprimirse en mayor medida. De hecho se puede demostrar que el algoritmo LZ alcanza el límite mínimo impuesto por el teorema de Shannon para cadenas muy largas. Es por tanto un algoritmo óptimo, al menos en teoría y para cadenas suficientemente largas. Aunque veremos a través de nuestros experimentos que puede encontrarse bastante lejos de ese límite incluso para ficheros tan grandes como los textos completos de la Biblia o del *Ulises*.

Shannon demostró que si se forma una cadena de n bit aleatorios y completamente independientes unos de otros, la entropía de cada uno de ellos es:

$$H = -p \log_2 p - (1-p) \log_2 (1-p)$$

en donde p es la probabilidad de que el bit sea un 1 y el logaritmo se toma en base 2. La entropía de la cadena es nH.

Veamos cómo se comporta el algoritmo LZ frente a este tipo de ficheros. Utilizando algún lenguaje de cálculo matemático (BASIC, C o FORTRAN pueden servir) se pueden crear fácilmente ficheros de conjuntos de bit aleatorios, que sean igual a uno con probabilidad p e igual a cero con probabilidad $1-p$. Yo he utilizado un programa comercial de cálculo matemático y he creado los ficheros *rand001.dat*, *rand01.dat*, etc., que contienen cien mil bit aleatorios y en los que la parte numérica indica el valor de p (0,01; 0,1; 0,2, etc.). Luego he comprimido todos los ficheros con un conocido compresor de Windows y he hecho una gráfica comparando el tamaño de los ficheros comprimidos con la fórmula de Shannon. Hay que tener la precaución de multiplicar los tamaños de los ficheros en bytes (caracteres ASCII) por 8, ya que un byte es una unidad de información que equivale a 8 bit. El resultado se puede ver en las figuras 1 y 2.



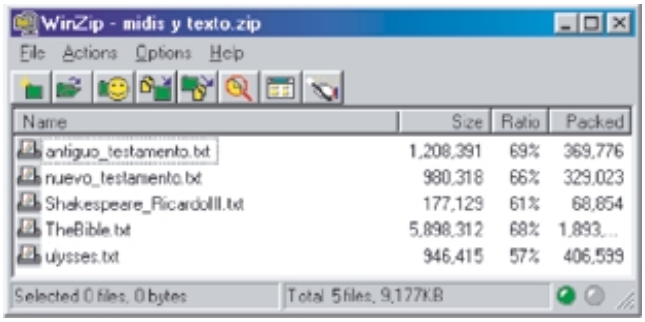
2. Una comparación entre el tamaño de los ficheros .zip *y el límite teórico impuesto por el teorema de Shannon*

El algoritmo LZ funciona bastante bien. Por supuesto que está siempre por encima del límite inferior impuesto por el teorema de Shannon, pero no se aleja mucho de él. Reproduce maravillosamente la forma de la curva, con lo que se puede apreciar que la fórmula de Shannon refleja una propiedad a primera vista tan intangible como la cantidad de información que contenga un fichero. La entropía de un fichero puede considerarse la información que contiene, puesto que es el tamaño mínimo al que puede reducírsele de forma reversible, es decir, de modo tal que pueda luego recuperarse intacto.

Vemos también que el compresor de ficheros nos proporciona una medida de la información o entropía de éstos, aunque esta medida pierde mucha de su precisión cuando se utilizan ficheros cuyos datos no sean independientes. Podemos verlo aplicándolo a ficheros de muy distinta procedencia, como he hecho en la figura 3, en donde se ve su efecto sobre ficheros de texto y de música. Como textos españoles he utilizado ficheros con los diez primeros libros del Antiguo Testamento y con el Nuevo Testamento completo, mientras que para el inglés se han utilizado la Biblia, *Ricardo III* de Shakespeare y el *Ulises* de James Joyce. Los textos bíblicos admiten una compresión que está en torno al 68 % con independencia del idioma en que se encuentren, porcentaje que es significativamente mayor que el 61 % de la compresión de Ricardo III. El Ulises, con un 57 %, es el que menor compresión admite, lo cual indica que contiene menos repeticiones y es más impredecible que los anteriores.

Con una compresión del 68 %, cada carácter ASCII (1 byte = 8 bit) ocupa $8 \times (1 - 0{,}68) = 2{,}56$ bit, mientras que ocupa 3,44 bit con otra del 57 %. Es interesante contrastar estos datos con la entropía del inglés, medida por el propio Shannon. Shannon demostró mediante sus célebres teoremas que la entropía de cualquier símbolo de una lista de símbolos independientes es:

$$H = -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3 - \ldots$$



3. Compresión de ficheros de texto

siendo p_1 la probabilidad de que se dé el primer símbolo, p_2 la probabilidad de que se dé el segundo, etc. La entropía del inglés puede calcularse utilizando esta fórmula si se conoce la probabilidad con la que aparece cada letra en un texto normal y estas probabilidades pueden obtenerse analizando textos. Se encuentra, por ejemplo, que la letra más probable es la E, que aparece un 13 % de las veces, mientras que las menos probables son la Q y la Z, que aparecen sólo un 0,1 %. Aplicando la fórmula de Shannon se obtiene que la entropía de cada letra de un texto inglés es 4,03 bit. Obsérvese que la entropía por carácter sería $\log_2 27 = 4{,}76$ bit en una lengua en la que las 26 letras del alfabeto y el espacio entre palabras aparecieran con la misma frecuencia.

Pero un texto inglés no es un conjunto de caracteres independientes entre sí. Por ejemplo, existen pares de letras más comunes que otros. El par TH es, en inglés, el más común de todos y aparece un 3,7 % de las veces, mientras que el par QT no aparece nunca. Esta correlación o dependencia entre letras reduce la incertidumbre o entropía del inglés con respecto a los 4,03 bit calculados suponiendo que las letras sean independientes. Para tener en cuenta la correlación se puede aplicar la fórmula de Shannon a pares de letras en lugar de a letras individuales. En tal caso los símbolos de la fórmula serán pares de letras, de los cuales existen $27 \times 27 = 6129$ (contando los espacios). La fórmula contiene entonces 6129 sumandos y su resultado es ligeramente inferior a 4,03.

Cuando en lugar de pares se consideran bloques de cuatro letras, la entropía se reduce a 2,8 bit por letra. Si como símbolos para la fórmula de Shannon se toman las palabras del diccionario inglés con sus respectivas probabilidades de aparecer en un texto, entonces la entropía que resulta es aproximadamente 9 bit por palabra, que equivale a 1,7 bit por letra. Puede continuarse este proceso considerando pares de palabras, y así sucesivamente. Obsérvese que las probabilidades de grupos de palabras dependen ya de aspectos gramaticales y semánticos de la lengua. Es decir, cuanto más amplio sea el bloque de letras o palabras con el que se calcula la entropía, más características de la lengua se recogen en dicho cálculo.

¿Puede calcularse la entropía del inglés teniendo en cuenta todas las posibles correlaciones, es decir, toda la estructura semántica y gramatical de la lengua? Shannon se dio cuenta de que es materialmente imposible registrar estadísticas de grupos amplios de palabras, mientras que los angloparlantes disponen para sus adentros de tal estadística. A partir de esta idea concibió un modo de medir la entropía del inglés haciendo que una persona tratara de adivinar determinada letra en un texto cuando se le presentaban las *n* letras que la preceden. Puede entonces codificarse la letra por el número de intentos necesarios para adivinarla y considerar la entropía del inglés igual a la entropía de esta serie de números. Con este procedimiento obtuvo entropías que decrecen cuanto mayor es el número *n* de letras precedentes que se le muestran al individuo. Cuando *n* es 3 la entropía es aproximadamente 3 bit, parecida a la calculada tomando probabilidades de grupos de cuatro letras. Pero continúa descendiendo hasta 1,3 bit cuando se presentan las 100 letras precedentes a la que hay que adivinar. Cuando *n* es mayor de cien, la entropía se mantiene prácticamente constante e igual a 1,3 bit, que puede considerarse por tanto la entropía del inglés. Esto indica también que la estructura de la lengua hace que estén relacionadas las letras situadas a una distancia de hasta 100 caracteres.

¿Cómo pueden interpretarse nuestros experimentos a la luz de estos resultados? Es claro que el compresor no puede detectar toda la estructura de la lengua a partir de un texto, aunque sea tan voluminoso como el Nuevo Testamento o el Ulises. Hemos visto que logra reducir el fichero a 3,44 bit por carácter en el caso del Ulises. Esta es aproximadamente la entropía del inglés cuando se consideran grupos de 3 letras. Por tanto, el compresor sólo ha sido capaz de "detectar" la estructura de la lengua en este nivel de grupos de 3 letras. En el caso de la Biblia consigue comprimir hasta 2,56 bit por carácter, que es la entropía cuando se consideran grupos de 5 letras.

Teniendo en cuenta que hay 19.683 posibles grupos de 3 caracteres, 531.441 de 4 caracteres y 14.348.907 grupos de 5 caracteres, está claro que el Ulises no tiene tamaño suficiente para lograr captar las frecuencias de los grupos de 4 letras, aunque aparezcan todos o una fracción apreciable de ellos. Tampoco la Biblia completa podría captar la estadística de los grupos de 5 caracteres aunque contuviera una proporción importante de ellos. Que el compresor de ficheros haya logrado reducir el tamaño hasta 2,56 nos indica que sólo una muy pequeña fracción de grupos de 5 caracteres aparece en el texto: para codificar uno de los grupos que aparecen necesitamos sólo $2{,}56 \times 5 = 12{,}8$ bit, que equivalen a $212{,}8 = 7131$ grupos, cifra muy inferior a los más de 14 millones de grupos posibles.

Naturalmente la "información" de que aquí estamos hablando es muy distinta de la habitual. Por exigencias de la compaginación de este número de la revista yo he tenido que "comprimir" bastante la contenida en la versión original del artículo y he tenido que echar mano de recursos muy distintos del algoritmo LZ. Pero el mes que viene seguiremos considerando otros aspectos interesantes de este asunto, junto con las ideas que aporten los lectores.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

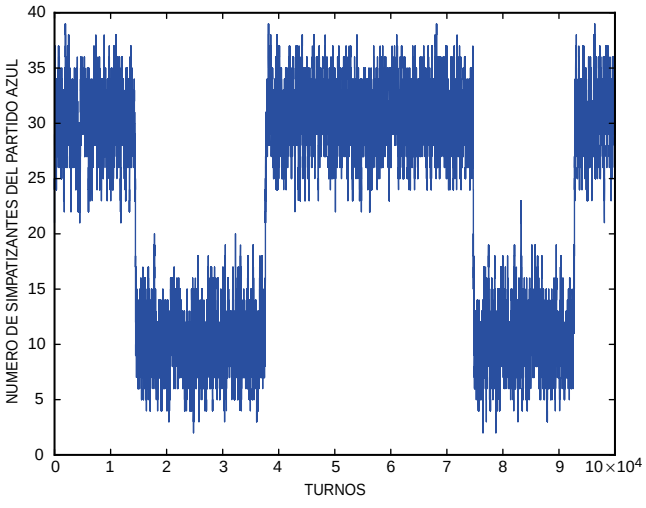
Las matemáticas de la opinión pública

¿Pueden tener algo que ver la opinión pública, los imanes y el origen de las fuerzas fundamentales de la naturaleza? Aunque se trate de asuntos tan dispares, los modelos matemáticos que los describen comparten una propiedad que ha sido fundamental en la física moderna: la *ruptura de simetría*. Afortunadamente, para ilustrar de modo básico esta propiedad no hace falta ser un experto en magnetismo o en teorías de unificación de campos de fuerza, sino que basta explorar un juego matemático que emula de modo extremadamente simplificado una lucha política completamente desprovista de ideología.

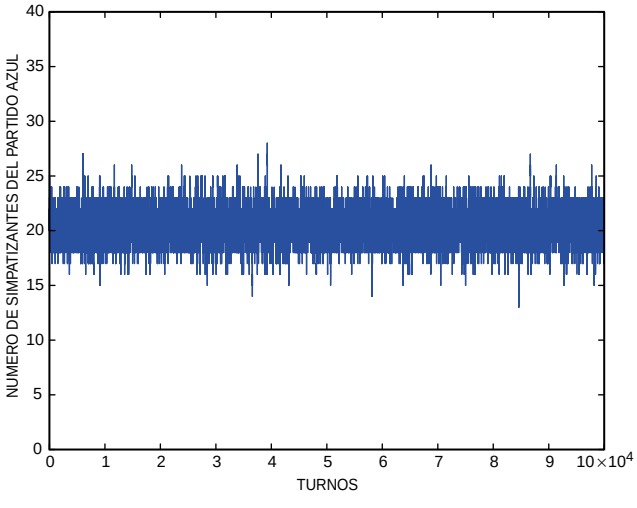
Dos grupos políticos, el Azul y el Blanco, se disputan el ayuntamiento de un pequeño pueblo de 41 habitantes. Cada uno de los habitantes se alinea con uno de los dos partidos, pero de vez en cuando puede cambiar de opinión. Podemos representar el "estado de opinión" de nuestro pueblo mediante un tablero con dos grandes casillas, una blanca y otra azul, en las que se encuentran 41 fichas. En cada turno se elige al azar una de las 41 fichas y se mueve de acuerdo con ciertas reglas. Imaginemos primero que a los habitantes de nuestro pueblo no les gusta demasiado discrepar, es decir, tienen una cierta tendencia a seguir los dictados de la mayoría. En este caso, las reglas para el movimiento de las fichas podrían ser las siguientes: la ficha elegida se coloca en la casilla en donde hay más fichas con una probabilidad 0,75 y se coloca en la casilla minoritaria con una probabilidad del 0,25.

Con estas sencillas reglas de juego, el comportamiento de la opinión pública del pueblo es ya bastante interesante. Supongamos que durante un gran número de turnos el partido Azul es el mayoritario. En toda esta fase del juego, el 75 % de las fichas movidas irá a parar a las filas azules y el 25 % a las blancas. Si la fase ha durado mucho más de 41 turnos, entonces el 75 % de la población, unas 30 personas, estará, en media, en el partido Azul y el 25 % restante en el Blanco. Sin embargo, esta distribución de la población sólo es cierta en media, ya que el comportamiento de cada individuo está sujeto al azar y habrá por ello fluctuaciones en torno a la distribución media. De hecho, las fluctuaciones pueden ser tan grandes como para inclinar la balanza hacia el partido contrincante.

En la gráfica de la figura 1 podemos ver cómo se comporta el número de simpatizantes del partido Azul. La he realizado simulando el juego en el ordenador durante 100.000 turnos. En los primeros momentos del juego, el partido Azul es mayoritario y el número de simpatizantes fluctúa en torno a 30. Hay fluctuaciones que hacen que el número se acerque a 20 y peligre la mayoría, pero logra recuperarla en varias ocasiones, hasta que, en un turno cercano al 15.000, el partido Blanco le arrebata la mayoría y la mantiene durante otros 23.000 turnos, aproximadamente. En ese período



1. Número de simpatizantes del partido Azul en función del número de turnos jugados, cuando los habitantes del pueblo tienen una probabilidad 0,75 de alinearse con el partido mayoritario

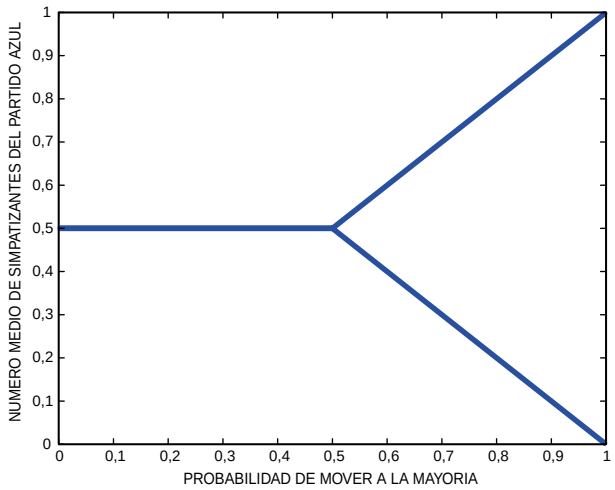


2. Número de simpatizantes del partido Azul en función del número de turnos jugados, cuando los habitantes del pueblo tienen una probabilidad 0,25 de alinearse con el partido mayoritario

el número de simpatizantes de partido Azul fluctúa en torno a 10 y el número de simpatizantes del Blanco fluctúa en torno a 30. Vemos que los cambios de mayoría se producen al azar, tras intervalos que son del orden de 20.000 turnos.

¿Qué ocurre si cambiamos la probabilidad de mover a la casilla mayoritaria? Siguiendo nuestro anterior razonamiento, se podría pensar que la fracción de habitantes del pueblo que simpatizan con el partido mayoritario es siempre igual a la probabilidad de mover a la casilla mayoritaria. Sin embargo, esto sólo es cierto si dicha probabilidad es mayor que 0,5. En el caso contrario nos encontramos ante individuos que prefieren alinearse con la opción minoritaria. Lo que ocurre entonces es que la proporción de simpatizantes de cada partido fluctúa en torno al 50 %. En efecto, en cuanto uno de los partidos comienza, por fluctuaciones, a tener más simpatizantes que el otro, los habitantes de nuestro pueblo, ahora amantes de las minorías, comenzarán a apoyar al partido minoritario. Es decir, ante cualquier desviación del reparto 50-50 entre los dos partidos, aparecerá una tendencia que tratará de restituir la situación de igualdad. Podemos verlo en la gráfica de la figura 2, en donde la probabilidad de que una ficha se mueva a la casilla mayoritaria es 0,25. Comprobamos también que las fluctuaciones ahora son bastante menos intensas que en el caso en que los habitantes preferían alinearse con el partido mayoritario (¿puede el lector averiguar por qué?).

Reuniendo todo lo que hemos deducido acerca del comportamiento de los habitantes de nuestro pueblo, podemos representar en una tercera gráfica el número medio de simpatizantes del partido Azul en función de la probabilidad de que una ficha se mueva a la casilla mayoritaria. El resultado se muestra en la figura 3. La línea que indica el número medio de simpatizantes se bifurca en dos cuando la probabilidad es 0,5; eso significa que, por encima de esa probabilidad, el número medio puede tomar dos valores posibles. El valor que de hecho toma depende del número inicial de simpatizantes de cada partido o, si éstas son cercanas al 50 % de la población total, dependerá de las fluctuaciones iniciales.



3. Número medio de simpatizantes del partido Azul en función de la probabilidad de alinearse con el partido mayoritario

Observe el lector que las reglas del juego no favorecen a ninguno de los dos partidos, pero estas mismas reglas, o más concretamente, el hecho de que dependan del comportamiento del resto de la población, hacen que uno de los partidos se convierta en mayoritario. En otras palabras, las reglas son simétricas ante el intercambio de partidos. Sin embargo, durante lapsos de tiempo bastante grandes (que aumentan considerablemente si se toma un pueblo con más habitantes), el "estado de opinión" no tiene esa simetría, sino que uno de los partidos vence a su oponente. Aunque ahora parezca simple, para la física fue una auténtica revolución percatarse de que el estado de un sistema puede no tener las simetrías que tienen las reglas que determinan el comportamiento de dicho sistema. Este es el fenómeno denominado *ruptura espontánea de simetría*. El adjetivo espontáneo es muy apropiado, ya que indica que la ruptura de la simetría se produce espontáneamente, debida a fluctuaciones, sin que haya nada en las reglas de evolución que indique hacia donde es realizará dicha ruptura.

En física, estas rupturas de simetría se comenzaron a estudiar en materiales magnéticos, es decir, en imanes. Un imán puede considerarse como un conjunto de pequeñas brújulas que tienden a orientarse en la dirección de un campo magnético y que, a su vez, crean campos magnéticos. De modo que cada brújula tiende a orientarse en la misma dirección que sus vecinas, igual que los habitantes de nuestro pueblo, en el caso en el que la probabilidad de adherirse a la mayoría sea mayor que 0,5, tienden a adoptar la opinión de la mayoría. En ausencia de campo magnético externo, las reglas que determinan la evolución de las pequeñas brújulas son perfectamente simétricas, ya que no favorecen ninguna dirección en especial. A pesar de ello, la tendencia de las brújulas a apuntar en la misma dirección que sus vecinas es capaz de romper esta simetría, siempre que la temperatura del imán sea suficientemente baja. Si la temperatura es alta, el comportamiento de las brújulas está sujeto a fuertes fluctuaciones, que son mayores que la propia tendencia a ordenar las orientaciones de las brújulas. Para temperaturas bajas, sin embargo, estamos ante una ruptura espontánea de simetría, que es la que hace posible que existan imanes permanentes como los que colocamos en la puerta de nuestro frigorífico. Una cuestión importante es que el estado macroscópico del imán, es decir, la orientación final de las brújulas, se determina a través de fluctuaciones que afectan a unas pocas brújulas, es decir, a fluctuaciones microscópicas. La ruptura de simetría puede considerarse como una amplificación de fluctuaciones microscópicas, como un mecanismo por el cual el azar microscópico invade el mundo macroscópico.

El concepto de ruptura de simetría ha sido utilizado también

en la teoría cuántica de campos, es decir, en el conjunto de teorías que tratan de explicar el origen de las fuerzas fundamentales de la naturaleza. En este caso, la simetría que se rompe es bastante complicada y se refiere a que las reglas de evolución de los campos de fuerza y de las partículas elementales no favorecen ninguna combinación de ciertas propiedades internas de las partículas y los campos, como la carga. La incorporación de la idea de ruptura espontánea de simetría, junto con algunos otros hallazgos, fue capaz de sacar a la teoría cuántica de campos de lo que muchos denominan la "era oscura" por la que atravesó en los años sesenta del siglo xx, y permitió establecer una teoría unificada de la fuerza electromagnética y la fuerza nuclear débil.

Volviendo a los modelos de comportamiento de la opinión pública, dejaré a los lectores la solución de un modelo un poco más complicado y que ya no tiene una relación estrecha con la ruptura de simetría. Este nuevo juego es una variación de un ejemplo creado por el matemático Ehud Kail, experto en teoría de juegos. En este caso la decisión que tienen que tomar los individuos de nuestro problema es si van o no a un cierto bar de moda en la ciudad. Otra novedad es que la población se divide ahora en tres grupos: los *conservadores*, los *poetas* y los *snobs*. Los conservadores tienden a tomar la decisión adoptada por la mayoría, mientras que los poetas quieren ser siempre singulares y prefieren alinearse con la minoría. Por último, los snobs quieren hacer lo que hacen los poetas, sin preocuparles si se trata de una decisión mayoritaria o minoritaria. El problema puede de nuevo interpretarse con un juego con dos casillas "ir al bar" y "quedarse en casa" y tres tipos de fichas. Las reglas podrían ser las siguientes: se elige una ficha al azar, si es conservador irá con una probabilidad 0,95 a la casilla mayoritaria; si es poeta, con una probabilidad del 0,05; si es snob, irá con una probabilidad del 0,95 a la casilla en donde haya mayoría de poetas. Se puede empezar escogiendo una población de 11 poetas, 31 snobs y 43 conservadores, pero invito a los lectores a explorar otras combinaciones, así como otras probabilidades de movimiento. ¿Cuál es, en media, la proporción de individuos en la casilla mayoritaria? ¿Cómo son las fluctuaciones en este juego? Lo discutiremos los próximos meses.

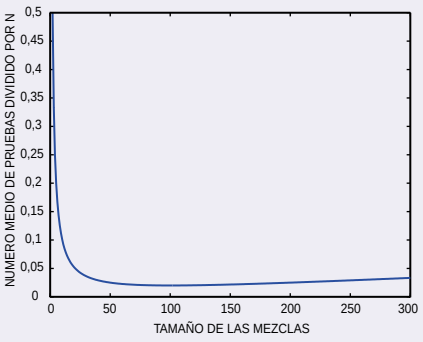
Solución del problema del banco de sangre

Jacobo Sánchez, estudiante de física de la Universidad de Valencia, me ha enviado una solución impecable del problema de las mezclas de sangre (Investigación y Ciencia del mes de agosto) para las pruebas del virus letal. Recordemos que se trataba de averiguar cuántas muestras de sangre se debían mezclar para que el número de pruebas a realizar fuera el mínimo. Expondré aquí una versión simplificada de la solución de Jacobo. El virus ataca a un 0,01 % de la población. Si se mezclan n muestras de sangre, la probabilidad de que la mezcla esté infectada es aproximadamente igual a 0,0001 por n. En este caso hay que realizar n pruebas para detectar cuáles de los integrantes de la muestra están infectados. Por otro lado, si la mezcla no está infectada, no hace falta hacer ninguna prueba adicional. Para cada mezcla, el número medio de pruebas a realizar por cada grupo mezclado (contando la que se realiza a la mezcla) es entonces $1 + 0{,}0001 \times n \times n = 1 + 0{,}0001 \times n^2$. Si tenemos un total de N muestras, al mezclarlas en grupos de n muestras, tendremos N/n grupos. Por tanto, el número total de pruebas, en media, es:

$$N/n\ (1 + 0{,}0001 \times n^2) = \\ = N\ (1/n + 0{,}0001 \times n)$$

Observen que esta fórmula tiene dos sumandos: el primero se hace más pequeño cuanto mayor es n, mientras que el segundo crece con n. Si se repasa el argumento que hemos utilizado, es fácil ver que el primer sumando proviene de la prueba que se hace a la mezcla, mientras que el segundo proviene de las pruebas que hay que hacer a las muestras individuales si el resultado de la prueba de la mezcla es positivo. Esta es la esencia del problema: cuantas más sangres se mezclan, menos grupos tenemos, pero por otro lado es más probable que la prueba de la mezcla sea positiva, en cuyo caso tenemos que realizar las pruebas a todas y cada una de las sangres que componen la muestra. El tamaño óptimo de la mezcla es aquel que logra un compromiso entre los dos sumandos de la fórmula.

La mejor forma de averiguar este tamaño óptimo es representar la cantidad $1/n + 0{,}0001 \times n$ en función del tamaño n de la mezcla. En la figura pueden ver la gráfica que se obtiene y comprobar que el mínimo se alcanza para n = 100. Por lo tanto, el banco de sangre debería mezclar 100 muestras de sangre. Con esta estrategia, un 1 % de las mezclas daría pruebas positivas y, por cada 100 muestras de sangre, habría que hacer, en media, sólo 2 pruebas, lo cual supone un ahorro considerable para el banco de sangre.



Número medio de pruebas por cada N muestras, en función del tamaño de las mezclas

Observemos también que la respuesta no coincide con la que da la teoría de la información, tal y como apuntábamos en el número de agosto. Con dicha estrategia la probabilidad de que la mezcla fuera positiva era 0,5, mientras que, con la calculada en los párrafos anteriores, es 0,01. Pero recordemos que aquella estrategia era óptima sólo si, en el caso de que la prueba de la mezcla fuera positiva, continuábamos el proceso realizando submezclas con las sangres sospechosas.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

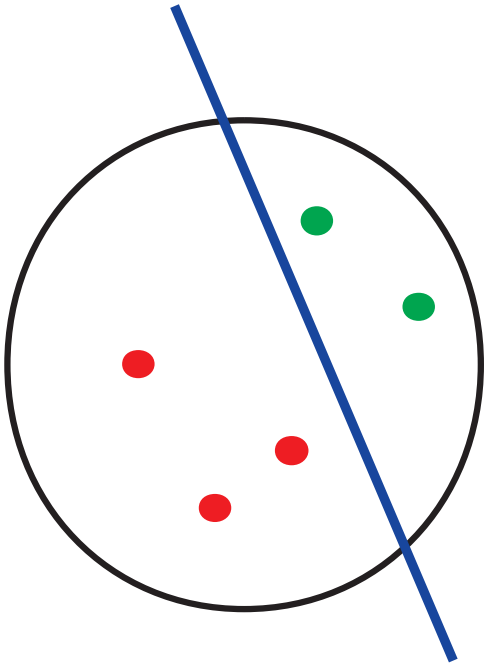
Las matemáticas del aprendizaje y la generalización

Anselmo Tajante es un tipo al que no le gustan las medias tintas. Para él las personas son simpáticas o antipáticas, inteligentes o necias, generosas o tacañas, diligentes o perezosas. El Sr. Tajante es, además, un tanto limitado: sólo dispone de estas cuatro categorías para clasificar a las personas que conoce. Para él, su amiga Alicia es simpática, inteligente, tacaña y perezosa, y su primo Bruno es antipático, inteligente, tacaño y diligente. De este modo, cada individuo puede ser codificado mediante un número con cuatro dígitos binarios, en donde el primero es un 1 si la persona es simpática y un 0 si es antipática; el segundo dígito es un 1 si la persona es inteligente y un 0 si es necia, y así sucesivamente. Alicia estaría codificada, para la reducida capacidad de apreciación de Anselmo, con el número 1100, y Bruno con el 0101. Como sólo existen $2^4 = 16$ números con cuatro dígitos binarios, Tajante es capaz de distinguir únicamente entre 16 categorías "psicológicas".

Esta sensibilidad tan tosca le facilita enormemente la generalización a partir de unos pocos ejemplos. Supongamos que Tajante se encuentra con Mister Dull, un inglés que dispone del mismo conjunto de categorías psicológicas. Para Dull no será muy difícil "enseñar" a Tajante el significado de los términos *lazy* (perezoso) y *diligent* (diligente) a través de ejemplos. Suponiendo que Dull conozca también a Alicia y a Bruno, y tenga la misma opinión de ellos, señalaría a ella como *lazy* y a él como *diligent*. Con estos datos, Tajante descartaría de su repertorio de categorías a los binomios inteligente-necio y generoso-tacaño. Con unos pocos ejemplos más, incluso tomados al azar, Tajante acabaría por deducir el significado de *lazy* y *diligent*.

El método de aprendizaje a través de ejemplos se utiliza ampliamente en el campo de la inteligencia artificial. Supongamos que queremos diseñar un programa de ordenador o un dispositivo que decida, a partir de una serie de datos, si una persona padece o no una determinada enfermedad. Una tarea como ésta puede resultar tan complicada, que sea prácticamente imposible diseñar un algoritmo que la lleve a cabo.

Una alternativa al diseño de algoritmos es utilizar un sistema capaz de realizar muchas clasificaciones y escoger una de ellas a través de ejemplos. Normalmente se emplea una red neuronal, que es un conjunto de pequeñas unidades de procesamiento conectadas entre sí. Dependiendo de la estructura y la intensidad de las conexiones, la red clasifica los datos de entrada de una u otra forma. No nos interesa ahora el funcionamiento detallado de una red neuronal. Lo único importante es que se trata de un dispositivo con un repertorio de clasificaciones y que podemos "enseñarla" a través de ejemplos: se toman los datos de una serie de individuos que se sabe si han contraído o no la enfermedad y se modifican las conexiones de la red de modo que los clasifique correctamente, esperando que así sea capaz de extraer las pautas que se esconden tras los ejemplos y clasifique correctamente a nuevos pacientes.



1. La secante define dos clasificaciones en el círculo: una que asigna 1 a los puntos verdes y 0 a los rojos y otra que asigna 0 a los verdes y 1 a los rojos

El procedimiento es idéntico al seguido por Dull para hacerle entender a Tajante el significado de *lazy*. Y, al igual que ocurría entonces, la rapidez con la que la red aprende es mayor cuanto menor sea su repertorio de clasificaciones. La diferencia con la historia de Dull y Tajante es que los dos compartían el mismo repertorio de clasificaciones, con lo cual el aprendizaje acaba siempre en la solución correcta. Por el contrario, en el caso de la red neuronal, la clasificación que queremos conseguir es parte del mundo real. Debemos, por tanto, asegurarnos de que el repertorio de clasificaciones de la red es suficientemente amplio como para contener la clasificación deseada, pero al mismo tiempo, suficientemente reducido como para que la red pueda "aprender" o "ge-

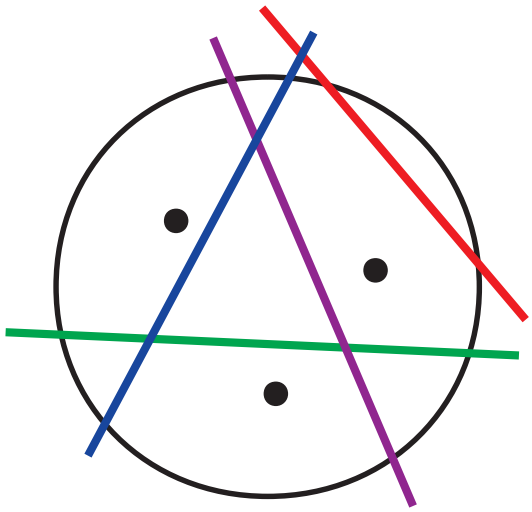
neralizar" a partir de un número limitado de ejemplos.

La teoría de la generalización estudia la capacidad de aprendizaje de un repertorio de clasificaciones a partir de ejemplos y a ella pertenece un concepto matemático fascinante por su universalidad: la dimensión de Vapnik-Chervonenkis (VC). La dimensión VC es un número que caracteriza la capacidad de generalización de cualquier repertorio de clasificaciones, aunque sea un repertorio infinito. Veamos con detalle en qué consiste.

Volvamos de nuevo a Tajante y al modo como clasifica a Alicia y a Bruno. Vimos que Alicia era simpática y Bruno antipático; así, con respecto a la clasificación simpático-antipático, a Alicia le corresponde un 1 y a Bruno un 0. A los posibles resultados de las clasificaciones que se pueden hacer con una serie de ejemplos se les llama dicotomías. La clasificación simpático-antipático en Alicia y Bruno crea la dicotomía 10. La clasificación inteligente-necio crea la dicotomía 11. Y las clasificaciones generoso-tacaño y perezoso-diligente crean, siempre en Alicia y Bruno, las dicotomías 00 y 01, respectivamente. El número de posibles dicotomías con dos personas es 4 y el repertorio de clasificaciones de que dispone Tajante es capaz de reproducirlas.

Pero si añadimos una tercera persona, el número de dicotomías posibles es $2^3 = 8$ y el repertorio de clasificaciones de Tajante es ahora incapaz de reproducirlas todas. Pues bien, la dimensión de Vapnik-Chervonenkis es el número de ejemplos a partir del cual el repertorio es incapaz de reproducir todas las dicotomías posibles. La dimensión VC de Tajante es, por tanto, 2.

Quizá le parezca al lector que el concepto es demasiado simple y que la dimensión VC es siempre el número n tal que 2^n es igual al número de clasificaciones disponibles en el repertorio. Sin embargo, no es así: la dimensión VC puede ser mucho menor que ese número n. Incluso un repertorio con infinitas clasificaciones puede tener una dimensión VC finita. Veamos un ejemplo. Los objetos a clasificar son los puntos de un círculo y las clasificaciones disponibles son las definidas por las secantes al círculo. En la figura 1 se nos ofrece una de esas clasificaciones y nos es dado observar cómo una secante define en realidad dos clasificaciones, una opuesta de la otra.



2. Las 8 dicotomías de tres puntos: la secante roja crea las dicotomías 000 y 111, la secante verde la 001 y la 110, la morada la 010 y la 101, y la azul la 100 y la 011

La dimensión VC de este repertorio de clasificaciones es 3, a pesar de que el número de secantes, y por tanto el número de clasificaciones, es infinito. La demostración no entraña especial dificultad. En la figura 2, pueden ver las ocho dicotomías sobre 3 puntos del círculo creadas por ocho clasificaciones definidas mediante cuatro secantes (recordemos que cada secante define dos clasificaciones). Por ejemplo, la secante azul define las dicotomías 100 y 011 (el orden de los puntos en las dicotomías es el siguiente: el primer punto es el superior izquierdo, el segundo el superior derecho y el tercero el inferior). El lector puede demostrar que es imposible encontrar cuatro puntos en el círculo y ocho secantes que reproduzcan las 16 dicotomías posibles. Por ello concluimos que la dimensión VC de las secantes es precisamente 3.

Como vemos en este ejemplo, la dimensión VC capta las limitaciones del repertorio de clasificaciones para clasificar puntos de modo arbitrario. A partir de cuatro puntos, las secantes son incapaces de reproducir las $2^4 = 16$ dicotomías posibles. Pero hay algo aún más fascinante en la dimensión VC. Como se deduce de la definición, el número máximo de dicotomías para l puntos es 2^l si l es menor que la dimensión VC. ¿Qué ocurre para un número de puntos mayor? Vapnik y Chervonenkis demostraron que, para cualquier repertorio de clasificaciones, el número máximo de dicotomías, para l mayor que la dimensión VC, depende de l como una potencia, más concretamente, como l elevado a la dimensión VC. Es decir, el número de dicotomías crece exponencialmente hasta la dimensión VC y, para un número de puntos superior, el crecimiento se reduce drásticamente (un crecimiento exponencial es siempre mucho mayor que uno en forma de potencia; basta comparar, por ejemplo, 2^{10} con 10 elevado a un número no muy grande). Lo más sorprendente es que este resultado es completamente general, válido para cualquier repertorio de clasificaciones, desde el que posee una red neuronal hasta el que pueda albergar su cerebro o el mío.

Vapnik y Chervonenkis demostraron también que el número de ejemplos necesario para que un dispositivo generalice, es decir, para que encuentre de forma unívoca la clasificación que corresponde a los ejemplos presentados, es del orden de la dimensión VC. Este resultado se aplica en la teoría del aprendizaje de redes neuronales y también en estadística matemática.

El cálculo de la dimensión VC es bastante difícil y, en la mayoría de los casos, sólo puede hacerse de modo aproximado. Pero la profundidad del concepto y la generalidad de los teoremas de Vapnik-Chervonenkis los convierten en uno de los logros más interesantes de la matemática de la segunda mitad del siglo XX.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Juegos cuánticos

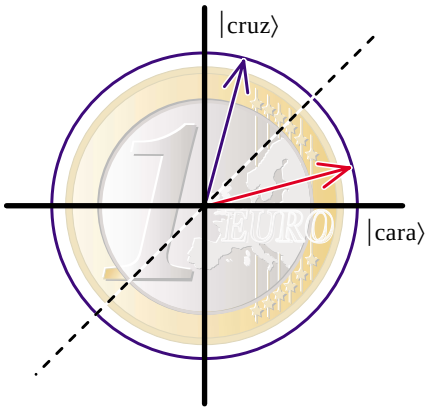
Supongo que la mayoría de los lectores de *Investigación y Ciencia* han oído hablar de la mecánica cuántica y conocen sus tres principales características: que se aplica a lo muy pequeño, que introduce el azar de forma ineludible en nuestra descripción de la naturaleza y que es sumamente extraña.

Una de las propiedades que dotan a la mecánica cuántica de ese carácter extraño es la llamada *dualidad onda-partícula*. Muchos experimentos de interferencia muestran que la luz es una onda; sin embargo, un fenómeno conocido como *efecto fotoeléctrico* no puede explicarse si no se admite que la luz tiene un comportamiento corpuscular, es decir, que se trata de un conjunto de partículas o corpúsculos. Por otro lado, las partículas que forman la materia (electrones, protones y neutrones), en la mayoría de las situaciones se comportan como corpúsculos, pero también pueden mostrar un comportamiento ondulatorio en ciertos experimentos de interferencia.

Sin embargo, ondas y partículas son cosas muy diferentes. Una onda consiste en ondulaciones de una cierta propiedad: el campo electromagnético, la presión del aire o la altura del agua en la superficie de un estanque. En estas ondulaciones la propiedad en cuestión puede tomar valores positivos y negativos en distintos puntos del espacio. Cuando una onda se encuentra con otra, los valores se suman, de modo que en los puntos en donde la primera onda es positiva y la segunda es negativa se produce una *interferencia destructiva*; esos puntos se comportan como si no existiera ninguna onda. Una partícula, por el contrario, no puede interferir destructivamente con otra: cuando se encuentra con una segunda partícula, lo único que puede hacer es chocar y volver a separarse o, si las dos partículas se atraen, formar una molécula o agregado.

¿Cómo es entonces posible que la luz, los electrones o los protones se comporten en ciertas ocasiones como ondas y en otras como partículas? ¿Cómo pueden reconciliarse en un mismo objeto estas dos naturalezas tan dispares? La mecánica cuántica resuelve el problema de la siguiente forma. El objeto —un electrón, un protón o un fotón— está descrito por la *función de onda*. La función de onda es una onda, es decir, ondulaciones de una determinada propiedad que puede tomar valores positivos o negativos. Pero esta propiedad no se parece a nada familiar: no es un campo eléctrico, ni una presión o una temperatura. De hecho, esta propiedad ni siquiera puede medirse directamente. Por ello algunos científicos y filósofos opinan que no corresponde a nada físico y que es un puro artificio matemático.

Es cierto que la función de onda no se parece a ninguna propiedad física conocida. Sin embargo, su *cuadrado*, es decir, el resultado de elevar al cuadrado el valor de la función de onda en cada punto del espacio, sí tiene un significado físico preciso: es la probabilidad de detectar al objeto en dicho punto. Si medimos la posición del objeto con una pantalla o un detector, sensible a fotones individuales, lo encontraremos en un solo punto, es decir, la pantalla obliga al objeto a manifestarse como partícula. El lugar en donde lo encontremos es aleatorio, pero la probabilidad de que el objeto aparezca en un punto dado es igual a la función de onda al cuadrado. Por ello, el valor de la función de onda en un punto se llama también *amplitud* de probabilidad. La mecánica cuántica parece demostrar que existe un azar ineludible en la naturaleza, un azar que no puede eliminarse con un mayor conocimiento del estado de un sistema. Pero este azar tiene además propiedades sorprendentes y muy distintas de las del azar al que estamos acostumbrados en el mundo macroscópico, sobre todo cuando se combinan varios sucesos.



Cómo cambia la función de onda de la moneda cuando Piccard le da la vuelta: la flecha roja indica el estado cuántico de la moneda antes de la acción de Piccard y la flecha azul el estado después de la misma

¿Cuál es la probabilidad de que al tirar dos dados salga un 5? El 5 puede obtenerse de varias formas: con un 1 en el primer dado y un 4 en el segundo; con un 4 en el primero y un 1 en el segundo; con un 2 y un 3 o, finalmente, con un 3 y un 2. Si los dados no están trucados, la probabilidad de cada una de estas cuatro posibilidades es igual a 1/36 (1/6 por 1/6). La probabilidad de que salga 5 es la probabilidad de que ocurra cualquiera de las cuatro posibilidades y se calcula simplemente sumando cada una de ellas: $1/36 + 1/36 + 1/36 + 1/36 = 1/9$. De este modo se trabaja con probabilidades: la probabilidad de que ocurra un suceso que puede darse de distintas formas (todas ellas incompatibles entre sí) es simplemente la suma de las probabi-

lidades de cada una de las formas posibles. El lector puede aplicar esta idea y demostrar, por ejemplo, que la probabilidad de que al tirar dos monedas salgan dos resultados iguales (cara-cara o cruz-cruz) es 1/4 + 1/4 = 1/2.

En mecánica cuántica esta sencilla regla no siempre es válida. Cuando un suceso cuántico puede ocurrir de varias formas posibles, no se suman las probabilidades, como en el caso "clásico", sino que se suman las funciones de onda o amplitudes; la probabilidad del suceso combinado se calcula como el cuadrado de esta amplitud suma. Esta nueva regla da lugar a fenómenos curiosos. Consideremos un suceso que puede darse de dos formas posibles y supongamos que la amplitud de la primera forma o posibilidad es 0,5 y la de la segunda es –0,5. Su suma es obviamente 0. La probabilidad de cada una de las formas posibles por separado es del 25 % (0,25 = 0,25) y, sin embargo, la probabilidad del suceso es 0. ¡Es como si en el problema de las dos monedas la probabilidad de que salga cara-cara fuera del 25 %, la de que salga cruz-cruz sea también del 25 %, pero fuera imposible que saliera cualquiera de las dos posibilidades! Esta afirmación parece completamente absurda y sin embargo así se comporta el mundo cuántico, por ejemplo, en un experimento de interferencia de fotones o electrones.

En la última década se ha comenzado a aplicar este insólito cálculo de probabilidades cuánticas en muchos otros ámbitos. Se habla de computación cuántica, criptografía cuántica y, recientemente, de juegos cuánticos. El primero de estos juegos fue propuesto hace dos años por David Meyer, matemático de la Universidad de California en San Diego. En él, el Capitán Piccard (personaje de la serie *Star Trek* cuyas iniciales coinciden con Classical Probability) juega contra Q (the Quantum), un individuo capaz de modificar el estado cuántico de un sistema.

Veamos primero la versión clásica del juego de Meyer. Se introduce una moneda en una caja con la cara hacia arriba y se juegan a continuación tres turnos: el primero a cargo de Q, el segundo de Piccard y el tercero de Q. En cada uno de ellos, el jugador puede, sin mirar la moneda y sin que el otro jugador vea lo que hace, darle la vuelta o dejarla como está. Después de los tres turnos se abre la caja y gana Q, si la moneda presenta una cruz y P, si presenta una cara. Es fácil ver que en este juego la probabilidad de ganar es siempre 1/2 para cada jugador si P o Q deciden aleatoriamente si dan la vuelta a la moneda o no.

Pero las cosas cambian si la moneda es cuántica y las probabilidades de que se muestre cara o cruz son en realidad el cuadrado de amplitudes de probabilidad. Si a y b son las amplitudes de las posibilidades "cara" y "cruz", respectivamente, representaremos la "función de onda" o estado cuántico de la moneda en la forma:

Utilizando la regla del cálculo

$$a|\text{cara}\rangle + b|\text{cruz}\rangle$$

cuántico de probabilidades, si la moneda está en este estado, la probabilidad de que veamos "cara" al abrir la caja será a^2 y de que veamos cruz, b^2. Por ello, las amplitudes tienen que verificar $a^2 + b^2 = 1$. Si representamos las amplitudes a y b como coordenadas en un plano, las posibles funciones de onda serán los puntos de un círculo de radio unidad (*el círculo azul en la figura*), ya que estos puntos verifican $a^2 + b^2 = 1$. Para ver mejor los puntos, representaremos la función de onda por un radio del círculo, como las flechas roja y azul que se muestran en la figura. El estado inicial de la moneda es o bien $|\text{cara}\rangle$ o bien $-|\text{cara}\rangle$ y, por tanto, será una flecha horizontal apuntando hacia la derecha o la izquierda.

P es un jugador "clásico", que únicamente puede dar la vuelta a la moneda o dejarla como está. Dar la vuelta a la moneda significa cambiar a por b. Un movimiento posible para Piccard es el que se muestra en la figura: la función de onda de la moneda es la flecha roja y Piccard decide voltearla; al hacerlo, el estado de la moneda pasa a ser la flecha azul. Supongamos ahora que Q puede manipular la moneda de forma no clásica. En la formulación original de Meyer, Q es capaz de girar las flechas a voluntad. Pero basta con que Q pueda hacer ciertos giros para que la ventaja frente a Piccard sea determinante. Por ejemplo, si Q puede girar la flecha 45 grados en el sentido contrario al de las agujas del reloj, entonces existe una estrategia que le dará siempre la victoria. En efecto, supongamos que el juego empieza con la flecha horizontal apuntando hacia la derecha. Lo que Q debe hacer es girarla 45 grados, de modo que el estado cuántico después de este primer turno será

Este estado tiene una curiosa pro-

$$1/\sqrt{2}|\text{cara}\rangle + 1/\sqrt{2}|\text{cruz}\rangle$$

piedad: si Piccard da la vuelta a la moneda, la función de onda de la misma no cambiará (es como si introdujéramos en la jaula del desdichado gato de Schrödinger un gas que mata a los gatos vivos y resucita a los muertos: el estado cuántico del gato no cambiaría). Por tanto, si en el tercer turno Q vuelve a hacer el giro, entonces, sea cual sea la acción tomada por Piccard en su turno, la flecha terminará apuntando hacia arriba. Abrimos la caja y encontramos, pues, que la moneda presenta una cruz: Q ha ganado.

El juego es bastante simple y ha recibido por ello algunas críticas que argumentan que en realidad no hay ningún misterio cuántico detrás de las ventajas de Q. Ekens, del Instituto de Tecnología de California, defiende esta idea proponiendo una versión puramente clásica del juego de Meyer.

Sin embargo, todos los investigadores están de acuerdo en que los jugadores cuánticos tienen considerables ventajas sobre los clásicos, ventajas que son la base de la excepcional potencia de los ordenadores cuánticos frente a los clásicos. Por tanto, la idea de Meyer puede contribuir a dilucidar cuál es la naturaleza de estas ventajas. Por mi parte, estoy convencido de que las probabilidades cuánticas pueden dar más "juego" y que en los próximos años surgirán más ejemplos e ideas acerca del tema.

parr-km0@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Información y juegos de azar: el problema de Monty Hall y la paradoja de los dos sobres

La forma en que se añade información a una situación incierta puede dar lugar a enigmas y paradojas. Un ejemplo ya clásico en la teoría de la probabilidad es el problema de Monty Hall. Otro no tan conocido, aunque no menos interesante, es la paradoja de los dos sobres, que presentaremos en este artículo pero a la que dedicaremos una exposición detallada en el próximo.

Monty Hall es el presentador de un viejo concurso de la televisión en Estados Unidos. En la fase final del concurso Monty enseña tres cofres a un sufrido concursante. En uno de ellos hay un gran premio y los otros dos están vacíos. El concursante elige nervioso uno de los cofres. Monty aparta entonces el cofre elegido y mira lenta y teatralmente en el interior de los otros dos. Cierra de nuevo uno de ellos, toma el otro con las dos manos y lo vuelca ante los ojos del concursante y el público: está vacío.

El concursante suspira aliviado, aunque no tiene ninguna razón para ello. Monty, generoso, le enseña los dos cofres que quedan cerrados y le ofrece la posibilidad de reconsiderar su decisión inicial: "Puede ahora escoger cualquiera de ellos", anuncia con un redoble de batería. ¿Qué debería hacer el concursante?

Mucha gente piensa que, una vez eliminado uno de los cofres, el premio puede estar por igual en los dos que quedan. Por lo tanto, no importa el cofre que se elija: la probabilidad de ganar el premio es del 50 %. Si a eso añadimos que, en la mayoría de la gente, modificar una decisión correcta produce una sensación bastante más dolorosa que mantenerse en una incorrecta, no es de extrañar que casi todo el mundo se niegue a cambiar de cofre. En una serie de charlas sobre probabilidad escenifiqué el concurso de Monty Hall y todos los "concursantes" sin excepción prefirieron quedarse con el cofre elegido en primer lugar.

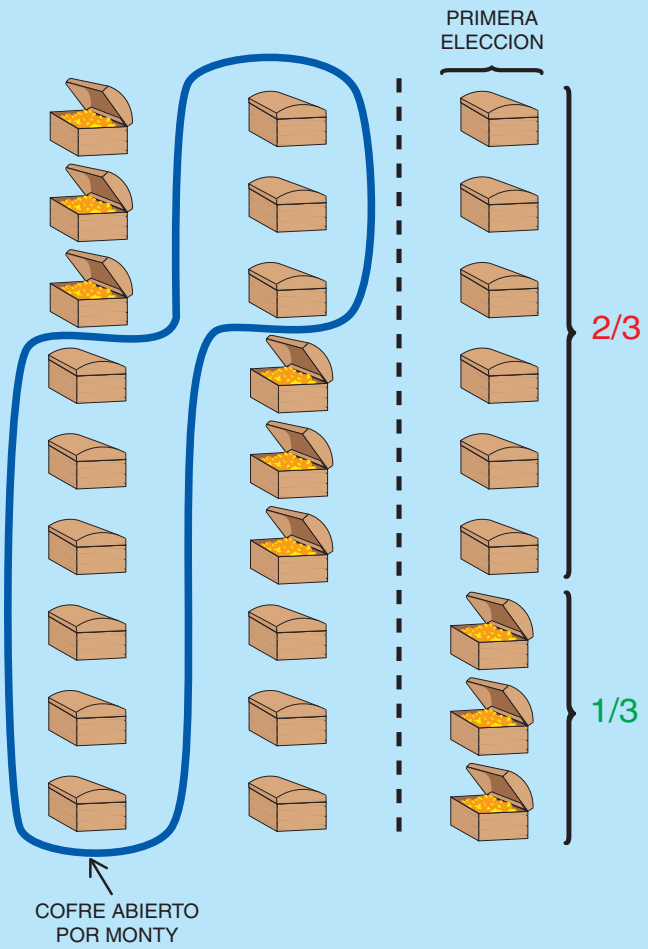
Sin embargo, lo mejor que puede hacer el concursante es cambiar su decisión inicial. Veamos por qué. Conviene primero que se imagine no un solo concurso, sino un gran número de ellos. Imagínese que el concursante tiene oportunidad de repetir el juego 900 veces, es decir, imagínese 900 *réplicas* del concurso, cada una con el premio en un cofre tomado al azar. Cuando el concursante elige por primera vez uno de los tres cofres, es evidente que aproximadamente un tercio de las veces acertará y dos tercios de las veces se equivocará. Es decir, *solamente* en un tercio de las réplicas, unas 300, el premio está en el cofre elegido por el concursante. Monty descubre el cofre vacío y quedan dos cofres cerrados. Recordemos que sólo en un tercio de las réplicas el premio está en el cofre elegido inicialmente. En el resto de las réplicas, el premio estará en el otro cofre.

Por tanto, si el concursante mantiene su decisión inicial, ganará un tercio de las veces. Si cambia, ganará dos tercios de las veces. Este argumento se muestra en la figura, en donde hemos supuesto 9 réplicas del juego. En términos de probabilidad, podemos decir que la probabilidad de ganar manteniendo la decisión inicial es un tercio, y la probabilidad de ganar al cambiar de cofre es de dos tercios. Los dos cofres no son equivalentes, como podía parecer a primera vista. Uno de ellos ha sido tomado al azar entre tres cofres en principio iguales (con igual probabilidad de contener el premio). El otro ha sido el superviviente de la inspección realizada por Monty.

El problema de Monty Hall nos enseña dos cosas interesantes acerca de la probabilidad y el azar. La primera es que, al introducir información en un sistema, cambian las probabilidades de los distintos sucesos que pueden ocurrir en dicho sistema. La segunda es más metodológica: el problema de Monty Hall es más difícil de entender si no se introducen las réplicas, es decir, si se piensa en un único concurso.

Veamos ahora la paradoja de los dos sobres, que tiene una formulación análoga al juego de Monty Hall, pero que da lugar a una situación bastante más sorprendente y cuya solución requiere adentrarse en conceptos más sutiles de la teoría de la probabilidad.

Ahora el presentador toma dos sobres e introduce en uno de ellos una cantidad de dinero x, desconocida para el concursante, y en el otro el doble de dicha cantidad, $2x$. El concursante elige uno de los dos sobre y lo abre. Supongamos que encuentra 1000 euros. El presentador entonces le ofrece al concursante la posibilidad de cambiar su elección original. ¿Cuál es la estrategia a seguir? En principio, los dos sobres parecen equivalentes y, por tanto, nadie creería que cambiar de sobre pueda suponer alguna ventaja. Sin embargo, en el sobre cerrado puede haber 2000 euros o 500 euros. Como no sabemos nada de las cantidades introducidas en los sobres, cada una de estas posibilidades se dará con una probabilidad 1/2. Por consiguiente, en el sobre cerrado habrá, en media, una cantidad 2000/2 + 500/2, es decir, 1250 euros. Si nos quedamos con el sobre abierto ganamos 1000 euros, pero si cam-



EL JUEGO DE MONTY HALL repetido nueve veces. En tres de ellas el concursante ha elegido el cofre premiado en su primera opción; en seis, el concursante elige un cofre sin premio. Después de que Monty abre uno de los cofres no elegidos (los rodeados por la línea azul), los dos restantes no son equivalentes: el elegido en primer lugar contiene el premio sólo un tercio de las veces, mientras que el que ha sobrevivido a la inspección de Monty contiene el premio dos tercios de las veces

biamos ganamos, en media, 1250 euros. La mejor estrategia será, pues, cambiar.

Pero este argumento conduce a una extraña paradoja. El argumento no depende de la cantidad hallada en el primer sobre (supongamos que en él hemos encontrado x euros; en el sobre cerrado habrá, en media, $2x/2 + (x/2)/2 = 5x/4$, que es mayor que x). Es decir, independientemente de lo que encontremos en el primer sobre, si cambiamos aumentaremos la ganancia media. ¿Para que abrir el primer sobre entonces? Antes de abrirlo ya sabemos que es mejor cambiar. Pero esta conclusión es completamente absurda, porque podemos aplicar de nuevo el argumento, que nos aconseja cambiar otra vez de sobre. Actuaríamos así como el asno de Buridán, alternando indefinidamente nuestras preferencias entre uno y otro sobre. La diferencia con el asno es que el argumento probabilístico que hemos descrito nos estaría diciendo que, cada vez que cambiamos nuestra preferencia de uno a otro sobre, sin necesidad de abrirlos, estamos aumentando la ganancia media. Hay que ser más necio que el asno para creerse semejante cosa.

Es evidente que en el argumento expuesto acerca del contenido del sobre cerrado hay una falacia. ¿Sabrá el lector encontrarla?

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

La paradoja de los dos sobres

El mes pasado abordábamos, junto con otros problemas de probabilidad, la paradoja de los dos sobres. El presentador de un concurso enseña dos sobres a un concursante. En uno de ellos ha introducido cierta cantidad de dinero y en el otro el doble de dicha cantidad. El concursante elige uno de los dos sobres, lo abre y comprueba que hay 100 euros. El presentador le ofrece ahora la posibilidad de cambiar de sobre. ¿Cuál es la mejor estrategia para el concursante?

Un argumento muy simple nos indica que el concursante debe cambiar de sobre. En efecto, en el sobre cerrado puede haber o bien 200 euros o bien 50. Si cada una de estas posibilidades es igual de probable, entonces el valor medio del dinero contenido en el sobre cerrado es $(200 + 50)/2 = 125$, que es superior a 100 euros. El argumento es válido cualquiera que sea la cantidad encontrada en el primer sobre. Si esa cantidad es x, en el segundo sobre puede haber $2x$ o $x/2$; si ambas posibilidades se dan con la misma probabilidad, el valor medio del dinero contenido en el segundo sobre será $(2x + x/2)/2 = 5x/4$, que es mayor que x. Por tanto, cambiar de sobre es siempre ventajoso. Resulta innecesario entonces mirar el contenido del primer sobre, ya que la decisión correcta es siempre cambiar. Pero esta conclusión es absurda, puesto que el mismo argumento se podría aplicar una y otra vez, sin abrir los sobres, y nos aconsejaría cambiar de uno a otro, con la disparatada pretensión de aumentar el valor medio de nuestra posible ganancia cada vez que cambiamos de sobre.

¿Dónde reside el error del argumento expuesto? El error consiste en suponer que las dos posibles cantidades para el segundo sobre, 200 o 50 euros, se dan con la misma probabilidad. En principio parece una suposición razonable, puesto que lo único que sabemos es que en uno de los sobres hay el doble de dinero que en el otro, pero no sabemos nada acerca de cómo se ha elegido la cantidad encerrada en cada uno de ellos.

Si, por ejemplo, sabemos que en el concurso no se van a utilizar céntimos de euro y obtenemos 99 € en el primer sobre, podemos inferir inmediatamente que en el segundo sobre habrá 198 €, ya que no puede contener 49,5 €. Cuando no conocemos cómo se han elegido las cantidades depositadas en los sobres, nada puede indicarnos si una de las dos posibilidades es más probable que la otra, de modo que lo más sensato es considerar que cada una tiene una probabilidad 1/2 de ocurrir. Sin embargo, se puede demostrar que la suposición es incorrecta, cualquiera que sea el procedimiento seguido para elegir las cantidades depositadas en los sobres. Vamos a ver primero un ejemplo que nos ayudará a entender el problema.

Imaginemos que las cantidades depositadas en los sobres se eligen de la siguiente forma: se toma al azar una cantidad entera (sin decimales) de euros entre 0 y 1000 €. Se introduce en un sobre dicha cantidad y en el otro el doble. Supongamos que el concursante conoce el procedimiento descrito. Si en el primer sobre encuentra una cantidad superior a 1000 €, es evidente que no debe cambiar de sobre. Si encuentra una cantidad x igual o inferior a 1000 € y *par*, entonces en el otro sobre puede haber o bien $2x$ o bien $x/2$; ambas posibilidades se dan con probabilidad 1/2. Se trata en este caso de una situación similar a la descrita al principio del artículo y la conclusión es la misma: el concursante debe cambiar de sobre.

Finalmente, si encuentra una cantidad inferior a 1000 € e *impar*, en el segundo sobre tiene que haber el doble de dicha cantidad; por tanto, lo más conveniente será cambiar.

Resumiendo: si el concursante encuentra en el primer sobre una cantidad superior a 1000 €, entonces no debe cambiar, mientras que si encuentra una cantidad igual o inferior a 1000 €, deberá cambiar de sobre. Como vemos, en el segundo sobre puede haber $2x$ o $x/2$, pero las dos posibilidades tienen la misma probabilidad sólo en el caso en que x sea igual o inferior a 1000 € y par. Si x es impar, entonces la posibilidad $2x$ tiene probabilidad 1 y la posibilidad $x/2$ tiene probabilidad nula. Por último, si x es mayor que 1000 € (y necesariamente par), entonces la posibilidad $x/2$ se da con probabilidad 1 y la posibilidad $2x$ no se da nunca.

¿Puede haber un procedimiento de elección de las cantidades en el que, para cualquier x, las dos posibilidades, $2x$ y $x/2$, se den con probabilidad 1/2? En un procedimiento de este tipo el 100 tendría que ser igual de probable que el 50 y el 200, y, a su vez, igual de probable que el 25 y el 400, y así sucesivamente. En otras palabras, todos los números de la secuencia infinita: ..., $x/8$, $x/4$, $x/2$, x, $2x$, $4x$, $8x$,... tendrían que aparecer con la misma probabilidad. Pero no existe un procedimiento capaz de extraer números de una secuencia infinita con igual probabilidad. Por eso, el argumento que dábamos al principio del artículo es incorrecto no sólo para ciertos procedimientos de elección de las cantidades, sino para cualquier procedimiento imaginable.

Pero, ¿qué ocurre cuando el concursante desconoce por completo el procedimiento seguido para elegir las cantidades de los sobres? ¿No hemos de suponer que las dos posibilidades, $2x$ y $x/2$, son igualmente probables? ¿Por qué habría de ser más probable una que otra? Lo que sabemos es que la probabilidad de cada una de esas posibilidades depende de la cantidad encontrada en el primer sobre, aunque el concursante no puede cal-

cular dicha probabilidad. No al menos en el primer intento.

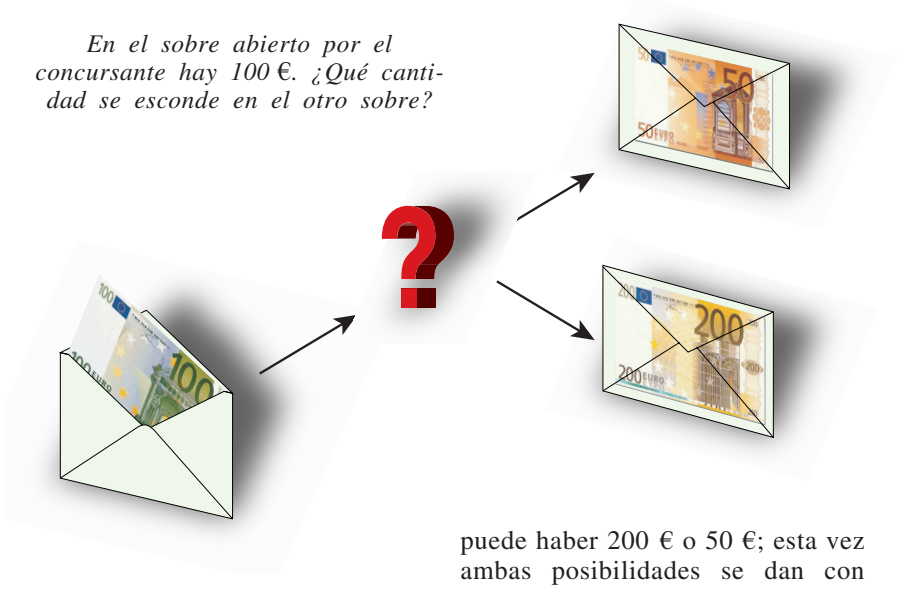
Si el juego se repite varias veces, el concursante puede utilizar las cantidades que salen en cada turno y estimar la probabilidad con la que aparecen, es decir, tratar de "deducir" el procedimiento seguido para elegir dichas cantidades. Al incorporar la información de los turnos jugados, el concursante puede elaborar un criterio de decisión similar al que hemos obtenido en el ejemplo anterior. Sin embargo, en su primer turno de juego, los dos sobres son equivalentes.

Otra objeción a nuestra crítica del argumento inicial podría ser la siguiente: ¿es realmente necesario un procedimiento para obtener las cantidades de cada sobre? ¿No se puede elegir en cada turno un procedimiento distinto? Esta objeción está relacionada con cuestiones básicas de la teoría de la probabilidad. Lo que hemos llamado hasta ahora "procedimiento" no es más que una cierta distribución de probabilidad para la cantidad depositada en uno de los dos sobres (la cantidad depositada en el otro sobre es simplemente el doble de la primera).

Que exista un "procedimiento" es lo mismo que decir que existe tal distribución de probabilidad. Esto equivale a que, si se repite el juego un gran número de veces, la frecuencia con la que aparecen las distintas cantidades posibles tiende a un cierto número al que llamamos probabilidad. Así ocurre, por ejemplo, cuando se lanza un dado no trucado un gran número de veces, pongamos 6000: cada uno de los seis números sale unas 1000 veces, aproximadamente, es decir, un sexto de las tiradas; cuantas más veces se lanza el dado, más se aproximan las frecuencias a este valor "ideal" o probabilidad, que es igual a un sexto. En nuestro caso, aunque eligiéramos al azar en cada turno el procedimiento por el cual se deciden las cantidades de los sobres, estos procedimientos al azar darían lugar a una cierta distribución de probabilidad, es decir, podrían considerarse como un único procedimiento.

Veamos, por último, una interesante modificación de la paradoja. En lugar de introducir en uno de

En el sobre abierto por el concursante hay 100 €. ¿Qué cantidad se esconde en el otro sobre?



los sobres el doble de dinero que en el otro, podemos utilizar otro tipo de operación matemática. Por ejemplo, en uno de los sobres depositamos una cantidad entera x de euros elegida al azar entre 0 y 1000 € y en el otro sobre introducimos una cantidad $f(x)$ definida de la siguiente forma: si x es menor o igual que 500 €, $f(x) = 2x$; si x es mayor que 500 €, $f(x) = 2x - 1001$ €.

El lector puede comprobar que la transformación que hemos definido asigna a cada entero entre 0 y 1000 uno y sólo un entero entre 0 y 1000. Decimos entonces que se trata de una transformación *biyectiva*, como también lo es la transformación inversa $f - 1(x)$. Gracias a ello, en este caso las dos posibilidades que surgen después de abrir el sobre elegido, es decir, que en el segundo sobre tengamos $f(x)$ o $f - 1(x)$, sí son igualmente probables. Sin embargo, la decisión de cambiar o no de sobre vuelve a depender de la cantidad encontrada en el primer sobre. Supongamos que encontramos 125 €. Entonces, en el segundo sobre puede haber 250 €, que es el resultado de aplicar la transformación $f(x)$ a 125, o 563 €, resultado, a su vez, de aplicar la transformación inversa $f - 1(x)$ a 125 (o, equivalentemente, 125 es el resultado de aplicar la transformación $f(x)$ a 563). Por tanto, en este caso, lo mejor es cambiar de sobre, ya que contiene una cantidad superior a la encontrada en el primer sobre.

Si la cantidad hallada en el primer sobre es 100 €, en el segundo puede haber 200 € o 50 €; esta vez ambas posibilidades se dan con probabilidad 1/2.

Nos encontramos de nuevo con la situación descrita en el argumento original de la paradoja. El valor medio del contenido del segundo sobre es 125 €, superior al contenido del primer sobre. Por ello, lo más sensato es cambiar. El lector puede comprobar que, si la cantidad hallada en el primer sobre es mayor que 500 €, entonces cambiar es una mala estrategia, a pesar de que en ciertos casos hay una cierta probabilidad de aumentar las ganancias. Por ejemplo, si en el primer sobre hay 625 €, en el segundo puede haber 249 € con probabilidad 1/2 o 813 € con la misma probabilidad. Sin embargo, el valor medio es 531 €, que es inferior a los 625 € que tenemos asegurados.

Para cualquier modificación del juego original, podemos obtener un criterio que nos diga si es mejor cambiar o no de sobre. Dicho criterio siempre dependerá de la cantidad encontrada en el primer sobre. De no ser así, tendríamos de nuevo la paradoja: se nos estaría diciendo que uno de los dos sobres es "mejor" que el otro, sin necesidad de abrir ninguno de ellos, lo cual es evidentemente absurdo.

En cualquier caso, la paradoja no deja de ser fascinante. Pese a su formulación simple, se requiere, para su correcta comprensión, considerar aspectos tan sutiles como la existencia de distribuciones de probabilidad sobre secuencias infinitas o incluso el propio significado del concepto de probabilidad.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Paradojas democráticas

En un congreso de teoría de juegos oí una charla sobre la votación para elegir la capital de la Alemania reunificada y cómo el resultado habría sido distinto si se hubiera utilizado otro método de votación. Más tarde, Raúl Toral, de la Universidad de las Islas Baleares, me envió un capítulo de un pequeño libro de Héctor Antoñana, *La Danza de los Números,* que trataba de la *paradoja de Condorcet*. Contaba historias curiosas sobre cómo el método de votación puede influir en el resultado. Algunas se referían a asuntos de cierta relevancia, como el caso de la capitalidad alemana, pero parecía que no eran más que casos singulares, poco útiles para reflexionar acerca de las limitaciones de la democracia.

Sin embargo, esta impresión ha cambiado al indagar más sobre el tema. Políticos y organizaciones, especialmente en Estados Unidos, están tratando de abrir el debate sobre los procedimientos de votación en las elecciones al senado y a los ayuntamientos, y defienden determinados métodos, como la llamada *votación aprobatoria* o el *método Condorcet*. Incluso ha habido proposiciones de ley al respecto, debatidas pero no aprobadas, en algunos estados.

De modo que lo que comenzó como un juego matemático ha resultado tener más trascendencia de lo que parecía inicialmente. Veamos cuál es la matemática que hay detrás de una simple votación.

Supongamos que una clase con 49 alumnos debe elegir a su delegado o representante. Hay tres candidatos, Inés, Diego y Carolina. Cada estudiante tiene sus propias predilecciones. Una suposición básica, y bastante razonable, de la teoría matemática de la votación afirma que si un estudiante prefiere a Inés antes que a Diego, y a Diego antes que a Carolina, entonces preferirá a Inés antes que a Carolina. Esta propiedad se llama *transitividad* y permite asociar a cada estudiante un orden de preferencia. En nuestro ejemplo, dicho orden sería Inés, Diego, Carolina. Lo indicaremos con la notación I>D>C. Supongamos que en nuestra clase 21 alumnos tienen como orden de preferencia I>>C>D, 3 alumnos tienen el orden C>I>D, 4 alumnos el C>D>I, 16 alumnos el D>C>I y, finalmente, 5 alumnos tienen el orden D>I>C (obsérvese que, en este ejemplo concreto, ningún alumno tiene el orden I>D>C). Estas preferencias están representadas en la figura.

Si se elige al delegado por medio de una votación única, y si cada alumno vota a su candidato preferido, Inés obtendrá 21 votos, Diego otros 21 (16 + 5) y Carolina sólo 7 votos (4 + 3). Para romper el empate entre Inés y Diego podemos hacer una nueva votación. Si todos los alumnos votan al candidato que prefieren entre los dos que quedan en esta segunda vuelta, entonces Diego ganará con 25 votos (4 + 16 + 5) frente a 24 (21 + 3) de Inés.

Sin embargo, cuando van a nombrar a Diego delegado de la clase, un seguidor de Carolina pide que levanten la mano los alumnos que prefieren a Carolina antes que a Diego. Para sorpresa de todos, 28 (21 + 3 + 4) alzan con decisión su mano, frente a sólo 21 (16 + 5) que prefieren a Diego. Cuando en la clase comienza extenderse la idea de que es absurdo tener como delegado a Diego cuando sólo una minoría lo prefiere a Carolina, y se está a punto de nombrar a ésta delegada de la clase, un seguidor de Inés pregunta de nuevo: ¿Quién prefiere a Inés antes que a Carolina? Se cuentan las manos levantadas y resulta que 26 (21 + 5) alumnos prefieren a Inés, frente a 23 (3 + 4 + + 16) que se quedarían con Carolina. La confusión se apodera de los alumnos.

Nuestra desorientada clase ha sido víctima de la llamada *paradoja de Condorcet*, en honor de Antoine de Caritat Condorcet, que estudió el problema a finales del siglo XVIII con la intención de encontrar el tamaño óptimo de los jurados que instauraría la revolución francesa. La paradoja advierte, en pocas palabras, que la transitividad de las preferencias de cada individuo no tiene por qué dar lugar a transitividad en las preferencias de un colectivo. En efecto, que una mayoría de la clase prefiera a Inés antes que a Carolina y a Carolina antes que a Diego no conduce necesariamente a que Inés sea preferida mayoritariamente sobre Diego. Se puede formar así un ciclo en las preferencias colectivas, como en el caso de nuestro ejemplo.

La paradoja de Condorcet no se produce siempre. Por ejemplo, si en nuestra clase eliminamos a los cinco votantes cuyas preferencias son D>I>C, entonces las votaciones "cara a cara" tendrían los siguientes resultados: Carolina obtendría 28 votos (21 + 3 + 4) frente a 16 de Diego. Inés, con 24 votos (21 + 3), también vencería a Diego con 20 (4 + 16). Y, finalmente, Carolina derrotaría a Inés con 23 votos (3 + 4 + 16) frente a 21. En este caso Carolina es preferida a los otros dos candidatos. Decimos entonces que Carolina es una *ganadora Condorcet*. La existencia de la paradoja de Condorcet es equivalente a que no exista ningún ganador Condorcet.

El sistema de enfrentar los candidatos por pares y quedarse con el ganador Condorcet es uno de los varios métodos posibles de elección de un candidato. Como hemos visto, de vez en cuando falla, puesto que ninguno de los candidatos es capaz de derrotar al resto. Cuando esto ocurre, el método de

votación de Condorcet se complementa con una regla simple. De los distintos enfrentamientos por pares se elimina el más reñido, es decir, aquél en el que la diferencia entre el ganador y el perdedor es mínima. Si eliminando este enfrentamiento aparece un ganador Condorcet, éste es el candidato elegido. Si no, se siguen eliminando enfrentamientos hasta que uno de los candidatos no es derrotado en ninguno de los enfrentamientos que quedan. Veamos cómo se aplica este método de eliminación a nuestro ejemplo. Los tres enfrentamientos entre nuestros candidatos son:

D > I: Diego gana a Inés por una diferencia de 1 voto.

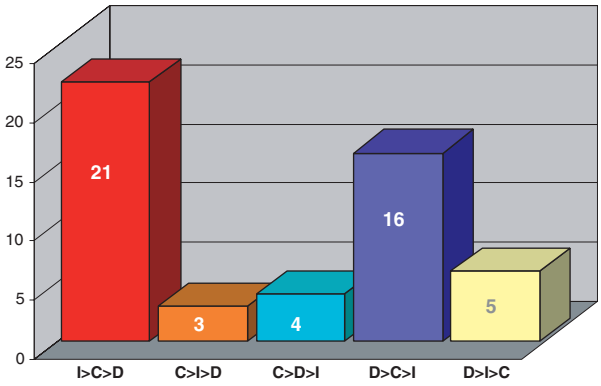
I > C: Inés gana a Carolina por una diferencia de 3 votos.

C > D: Carolina gana a Diego por una diferencia de 7 votos.

Al eliminar el enfrentamiento más reñido, D > I, Inés resulta ser la ganadora Condorcet, porque nunca es derrotada en los dos enfrentamientos que quedan. Es decir, el método de Condorcet da a Inés como ganadora. Sin embargo, en una votación simple entre los tres candidatos, con su correspondiente segunda vuelta de desempate, Diego es el candidato elegido.

Existen otros métodos de votación que a priori parecen igual de razonables. El llamado *recuento de Borda* consiste en que cada votante puntúa a los candidatos de acuerdo con su preferencia: con 2 puntos al primero, 1 punto al segundo y 0 puntos al tercero. Si utilizamos este método en nuestra clase, Inés obtiene $21 \times 2 + 3 + 5 = 50$ puntos. La puntuación de Carolina es $21 + 3 \times 2 + 4 \times 2 + 16 = 51$ puntos y la de Diego $4 + 16 \times 2 + 5 \times 2 = 46$ puntos. Por tanto, con el método de votación de Borda la ganadora es Carolina.

Otra posibilidad es la *votación aprobatoria*, en la que cada votante escribe en la papeleta el nombre de todos los candidatos "que no le desagradan", es decir, los candidatos que aprueba y que desearía que fueran elegidos. Se cuenta el número de veces que aparece el nombre de cada candidato (sin importar el orden en que aparece en cada papeleta) y gana el que ha sido aprobado por más votantes. Para encontrar el candidato que elegiría nuestra clase con este método es necesario hacer algunas suposiciones adicionales. Imaginemos que todos los votantes aprueban a los dos primeros candidatos de su orden particular. En ese caso Inés sería aprobada por $21 + 3 + 5 = 29$ votantes, Carolina por $21 + 3 + 4 + 16 = 44$, y Diego por $4 + 16 + 5 = 25$. Carolina, por tanto, sería la ganadora.



Número de votantes de cada orden de preferencia

Si dependiendo del método de votación puede salir elegido cualquiera de los tres candidatos, ¿cuál es entonces el más justo? Las matemáticas no pueden dar una respuesta nítida a esta pregunta. De hecho, lo que las matemáticas han demostrado es que no existe ningún método de votación "perfecto". Este resultado se conoce como teorema de Arrow. Keneth J. Arrow, profesor de Stanford y premio Nobel de Economía, estableció una serie de condiciones muy básicas que cualquier método de votación debería satisfacer, para luego demostrar que dichas condiciones son incompatibles, es decir, que no existe ningún método que las cumpla en su totalidad. Como decía, las condiciones son muy básicas. Una de ellas, por ejemplo, exige que al eliminar un candidato de una votación el resultado no varíe, a no ser que el candidato eliminado sea el ganador. Esta regla tan simple no se cumple en el método de Condorcet. Sin embargo, los otros métodos también violan condiciones igual de razonables.

La elección del método de votación no es, por tanto, sencilla; se debe hacer utilizando argumentos que van más allá de las matemáticas. Por ejemplo, existen organizaciones políticas que defienden la implantación de la votación aprobatoria destacando las ventajas de este método: con él, el llamado voto útil o voto estratégico carece de sentido, puesto que el votante puede aprobar a tantos candidatos cuantos desee. Por otro lado, las campañas electorales serían menos agresivas y los candidatos se preocuparían más de defender sus programas en lugar de atacar los de los contrincantes. De hecho, la votación aprobatoria se usa en varias sociedades científicas para elegir a los miembros de su junta directiva. Existen multitud de páginas en Internet explicando y defendiendo diferentes métodos de votación. Un sitio interesante es electionmethods.org, que se decanta por el método de Condorcet.

La teoría matemática de las votaciones es mucho más amplia de lo que hemos comentado aquí. Estudia las votaciones realizadas por fases, así como la importancia en ellas de los votos estratégicos, es decir, aquellos en los que el votante no vota de acuerdo con su orden de preferencia en alguna de las fases con el objetivo de que su candidato preferido salga finalmente elegido. Analiza también la posibilidad de crear coaliciones de votantes en votaciones ponderadas, además de otras situaciones interesantes.

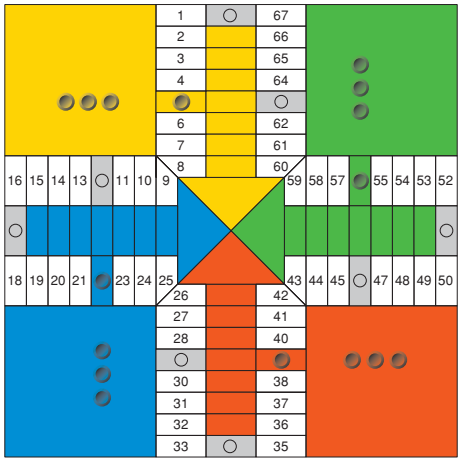
Como hemos visto, el problema de extraer una decisión colectiva a partir de las preferencias de los individuos que forman un grupo no es tan simple como puede parecer en un principio. Admite distintas soluciones que dan lugar a diferentes resultados; no es fácil decidir cuál de ellas es la más justa o la que mejor representa la opinión del grupo. En adelante, cuando nos digan aquello de que "la democracia es el menos malo de los sistemas políticos", podremos replicar: "De acuerdo, pero, ¿qué democracia?".

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Ventajas engañosas

¿Qué diría usted si en un juego de azar le ofrecieran aumentar las probabilidades de ganar en cada turno? Probablemente pensaría que su oponente había perdido el juicio. Sin embargo, en ocasiones, lo que en principio parece ventajoso puede ser incluso desfavorable. Recordemos, por ejemplo, las reglas del juego del parchís. Sacar un seis en el dado es muy ventajoso, puesto que avanzamos el mayor número de casillas y además repetimos tirada. Existe, eso sí, una regla, que podríamos llamar "de compensación", según la cual si se saca seis tres veces consecutivas la última ficha movida vuelve a casa. Si alguien nos ofrece jugar al parchís con un dado trucado, en el que el seis tiene una probabilidad de salir mayor de lo habitual, ¿deberíamos aceptar la oferta, o será preferible jugar con un dado normal? Responder a esta pregunta es bastante complicado. Más adelante haremos un análisis parcial del problema. Veamos antes un juego más simple en el que aumentar las probabilidades de ganar en cada turno da lugar a una disminución de las ganancias.



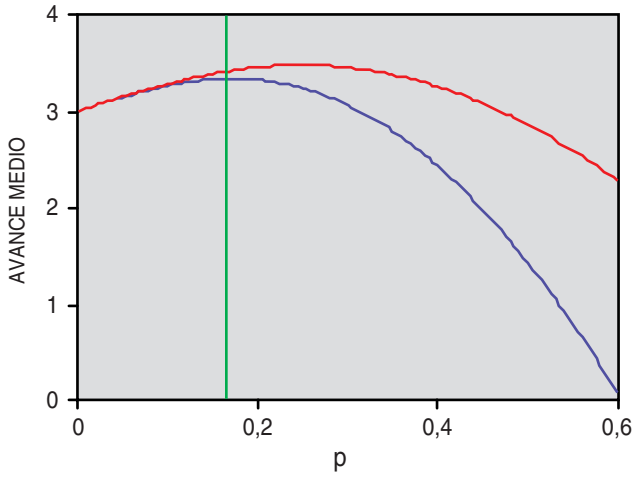
1. En el parchís tres seises nos "devuelven a casa". ¿Merece la pena jugar con un dado trucado en el que el seis sale con mayor probabilidad?

Christian van den Broeck y Bart Cleuren, físicos del Centro Universitario de Limburg, en Bélgica, estudian juegos de este tipo y sistemas físicos y químicos relacionados con ellos, como un conjunto de partículas que, al ser empujadas en un sentido, se mueven en el contrario. Ellos las llaman *donkey particles*, es decir, "partículas burro", aludiendo a la costumbre que estos tozudos animales tienen de moverse siempre en contra de la fuerza que se ejerce sobre ellos. Por extensión, los juegos en los que las ganancias disminuyen cuando se aumentan las probabilidades de ganar en cada turno se llaman *donkey games* o "juegos burro".

Veamos el más simple de esos juegos burro, estrechamente relacionado con la regla del parchís que hemos mencionado antes. Se juega con una moneda en la que sale CARA con probabilidad p y CRUZ con probabilidad $1-p$. Si al lanzar la moneda sale CARA, entonces el jugador gana un euro. Si sale CRUZ, lo pierde. Pero si sale el mismo resultado dos veces seguidas, se cancela la ganancia o pérdida anterior. Por ejemplo, si obtenemos CRUZ-CARA-CARA-CARA la ganancia final será cero, ya que primero perdemos un euro, luego lo ganamos, en el tercer turno esa ganancia se cancela y, finalmente, en el cuarto volvemos a ganar un euro.

Si la moneda no está trucada, es decir, si p es 1/2, el jugador no tiene ninguna ventaja ni desventaja: el juego es justo y no hay una mayor tendencia a ganar o a perder. Pero ¿qué ocurre si modificamos p? ¿Qué es más beneficioso para el jugador, que p aumente o que disminuya? Un modo de encontrar la respuesta es realizar una simulación por ordenador del juego para distintos valores de la probabilidad p. Pero también es posible encontrar la solución mediante un argumento matemático. Para ello se requieren algunos conocimientos básicos de matemáticas y un poco de reflexión. En el recuadro se describe este argumento para los lectores interesados. El resultado final se puede ver en la gráfica, en la que se muestra la ganancia media en cada turno en función de p. Como cabría esperar, el juego es justo si $p = 1/2$. Lo interesante es que, para p entre 0 y 1/2, la ganancia media es positiva, es decir, el juego es ganador, mientras que para p entre 1/2 y 1 es perdedor. Es decir, si comenzamos con el juego justo, $p = 1/2$, y aumentamos la probabilidad p de ganar en cada turno, convertimos al juego en perdedor, mientras que si disminuimos la probabilidad de ganar en cada turno, hacemos que el juego sea ganador. Tenemos por tanto un claro ejemplo de "juego burro", en el que aumentar p es una "ventaja engañosa".

Volvamos a la pregunta de partida acerca de las ventajas de un dado trucado en el parchís. Hacer un análisis completo del juego es muy complicado, ya que habría que tener en cuenta la probabilidad de obtener un 5 para sacar fichas de casa, y detalles de este tipo. Para encontrar una respuesta al menos aproximada he simplificado bastante el problema. Considero un juego parecido al de Van den Broeck y Cleuren: se lanza un dado en el que la probabilidad de que salga un seis es p y la probabilidad de que salga otro número, del uno al cinco, es $(1-p)/5$. Una ficha avanza tantas casillas como marque el dado. Cuando sale un seis tres veces seguidas la ficha retrocede 36 casillas (el número de casillas de "casa" a "meta" en el parchís es 72, por eso la posición media de una ficha será, aproximadamente, 36). Al calcular el avance medio en una tirada resulta la



2. Avance medio en cada tirada con un dado trucado (p es la probablidad de sacar un seis) en un "parchís simplificado"

curva azul de la figura 2. Si, en lugar de penalizar con 36 casillas los tres seises consecutivos, hacemos que la ficha retroceda 16 casillas, el resultado será el que se muestra en la curva roja. La línea vertical verde indica $p = 1/6$, es decir, el caso de un dado no trucado. Curiosamente, el máximo de la curva azul está muy cerca de este valor $p = 1/6$. Esto nos indica que cualquier desviación con respecto al dado no trucado presenta una desventaja para el jugador, y además esta desventaja es mínima. ¿Es esto una coincidencia o la regla de "volver a casa con tres seises" se diseñó precisamente para cancelar las posibles ventajas o desventajas producidas por dados imperfectos? Me temo que no, puesto que nuestro análisis es bastante simplificado y un estudio más cuidadoso probablemente daría lugar a una gráfica más parecida a la curva roja que a la azul.

Existe un problema clásico relacionado con esta cuestión: ¿cómo puede jugarse con una moneda, de la que se sabe que está trucada pero no cuál es la probabilidad p de que salga CARA, de manera que el jugador pierda o gane en cada turno con una probabilidad ½? Recuerde que no conocemos p; por lo tanto, la solución debe consistir en un procedimiento que nos dé, sea cual sea el valor de p, dos posibles resultados perfectamente equiprobables. La moneda se puede lanzar varias veces por turno. La solución es muy simple y evidente, pero encontrarla no es nada fácil y requiere una considerable dosis de "inspiración". La revelaremos en el próximo número.

El "juego burro" más simple

Supongamos que en el "juego burro" de la moneda anotamos en una pizarra el resultado de las jugadas de cada turno de acuerdo con las reglas siguientes: si la pizarra está vacía, escribimos el resultado del lanzamiento de la moneda; si sale el mismo resultado que está apuntado en la pizarra, la borramos y la dejamos vacía; si sale un resultado distinto, la borramos y escribimos el que ha salido. Es decir, en la pizarra anotamos los últimos resultados que están "pendientes" de consolidarse o de cancelarse. Las reglas establecen que el pago que recibe el jugador en un turno depende de lo que resulte del lanzamiento de la moneda y de lo que esté anotado en la pizarra. La siguiente tabla detalla todas las posibilidades. Hemos indicado también entre paréntesis cómo queda la pizarra tras el turno:

		VALOR ANOTADO EN LA PIZARRA ANTES DE LANZAR		
		NINGUNO	CARA	CRUZ
RESULTADO DE LA TIRADA	CARA	+1 (CARA)	–1 (NINGUNO)	+1 (CARA)
	CRUZ	–1 (CRUZ)	–1 (CRUZ)	+1 (NINGUNO)

En los problemas de probabilidades siempre conviene imaginarse muchos jugadores, cada uno de ellos jugando simultáneamente pero de modo independiente, y cada uno con su moneda y su pizarra. Si tenemos N_0 jugadores en cuya pizarra no hay nada, N_1 jugadores con el valor CARA anotado en ella y N_2 jugadores con el valor CRUZ, ¿qué valores tendrán anotados en el turno siguiente? De los N_0 jugadores con la pizarra vacía, una fracción p sacará CARA y una fracción 1 – p sacará CRUZ en el siguiente turno (recordemos que p es la probabilidad de que, al lanzar las monedas, el resultado sea CARA). En ambos casos, estos jugadores anotarán el valor obtenido en su pizarra correspondiente. A su vez, de los N_1 jugadores con CARA en su pizarra, una fracción p sacará CARA en el siguiente turno, y en consecuencia borrará lo escrito, mientras que una fracción 1 – p sacará CRUZ y anotará dicho resultado. Por tanto, en el turno siguiente el número de jugadores con la palabra CRUZ escrita en su pizarra será:

$$N_2' = (1-p)N_0 + (1-p)N_1 \qquad (1)$$

Con un argumento idéntico, se puede obtener el número de jugadores con la pizarra vacía:

$$N_0' = pN_1 + (1-p)N_2 \qquad (2)$$

y el número de jugadores con la palabra CARA:

$$N_1' = pN_0 + pN_2 \qquad (3)$$

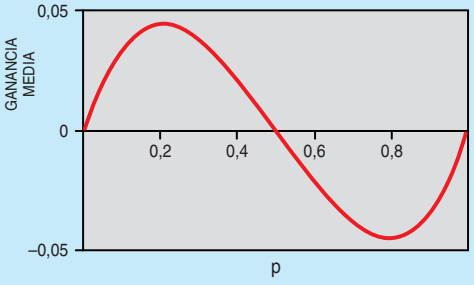
Por otro lado, si conocemos N_0, N_1 y N_2 podremos también saber cuál es la ganancia neta de nuestro conjunto de jugadores. Por ejemplo, los N_1 jugadores en cuyas pizarras está anotado el valor CARA pierden siempre un euro. Lo ganan los N_2 jugadores con CRUZ en sus pizarras. Finalmente, de los N_0 jugadores con la pizarra vacía, pN_0 ganan un euro y $(1-p)N_0$ lo pierden. Por lo tanto, la ganancia neta en un turno es:

$$G = pN_0 - (1-p)N_0 - N_1 + N_2 \qquad (4)$$

Cuando se ha jugado un número considerable de turnos, ocurre que N_0, N_1 y N_2 alcanzan unos valores constantes, que también se llaman estacionarios. Son valores que tienen la peculiaridad siguiente: si se introducen en las ecuaciones (1), (2) y (3), los resultados de dichas ecuaciones, es decir, N_0', N_1' y N_2', coinciden exactamente con N_0, N_1 y N_2. Esta peculiaridad nos permite calcular estos valores estacionarios y, introduciéndolos en la ecuación (4), obtener la ganancia neta. El lector interesado podrá comprobar que la ganancia neta por jugador resulta ser:

$$g = \frac{G}{N_0 + N_1 + N_2} = \frac{p(1-p)(1-2p)}{(2-p)(1+p)}$$

tal y como se representa en la gráfica.



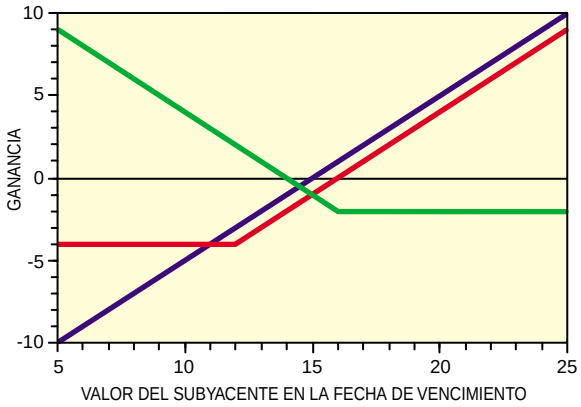
Ganancia media en cada turno del "juego burro" (p es la probabilidad de que salga cara en la moneda trucada)

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Jugar con opciones y futuros

En la bolsa ya no sólo se compran y venden acciones, es decir, participaciones en el capital de las empresas, sino que, desde hace unas cuantas décadas, se intercambian además *productos derivados*, de los cuales los más conocidos son probablemente los contratos de futuros y de opciones. Este mes vamos a presentar un juego que reproduce este mercado. Se debe a David Epstein, un alumno del conocido experto en análisis financieros Paul Wilmott. Pero veamos antes en qué consisten esos productos derivados.



1. Ganancia para el comprador de diferentes contratos en función del valor del subyacente: en azul, un futuro a 15, en rojo una opción de compra (call) a 12 y con una prima de 4, y en verde una opción de venta (put) a 16 con una prima de 2

Un contrato de futuro es un acuerdo entre dos individuos por el cual uno de ellos (el comprador) se compromete a comprar algo en una determinada fecha futura al otro individuo (el vendedor) y a un precio pactado en el contrato, que se llama *precio de futuro*. Es decir, un contrato de futuro no es más que una venta aplazada, en la que el precio se establece en el momento de realizar el contrato pero la venta no se completa hasta la fecha de vencimiento. Lo que se va a comprar o vender se llama *activo subyacente*, y puede tratarse de acciones que cotizan en bolsa, divisas, oro o incluso naranjas. Siempre se trata de cosas que tienen un precio en algún mercado, ya sea en la bolsa, en el mercado de divisas o en el de cítricos, pero este precio sufre variaciones en principio impredecibles. El futuro es por tanto una apuesta: si en la fecha de vencimiento, el precio del subyacente en el mercado es inferior al precio de futuro, entonces el comprador pierde, porque está obligado a comprar a un precio superior al del mercado. Si el precio del subyacente es alto y supera al precio de futuro, entonces el que gana es el comprador. En la figura 1, la curva azul representa la ganancia del comprador de un futuro a 15 en función del precio del subyacente. La ganancia del vendedor es la misma cambiada de signo.

Los contratos de opciones pueden ser de dos tipos: opciones de compra (*call*) y opciones de venta (*put*). Una opción de compra es un contrato por el cual un individuo (el comprador de la opción) adquiere la opción de comprarle algo (el activo subyacente) a otro (el vendedor de la opción) por un precio y en una determinada fecha acordados en el contrato. Ahora el precio pactado se llama *precio de opción* y, a diferencia de los futuros, el comprador debe pagar al vendedor una cierta cantidad, una *prima* que se pacta entre los dos contratantes. En este caso, en la fecha de vencimiento el comprador no tiene obligación de comprar el subyacente. Por lo tanto, si el precio de mercado del subyacente es inferior al precio de opción, el comprador de la opción no la ejecutará y habrá perdido la prima. Pero si el precio del subyacente es mayor que el precio de opción, entonces el comprador ejercerá la opción de compra y su ganancia en la transacción será la diferencia entre el valor del subyacente y el precio de opción. La ganancia total será entonces el valor del subyacente menos el precio de opción menos la prima, tal y como se muestra en la curva roja de figura 1 para un ejemplo de opción a 12 con una prima de 4.

En la opción de venta (*put*) ocurre al contrario. El comprador de la opción adquiere, a cambio de la prima, la opción de vender el subyacente. La ganancia del comprador de la opción es ahora mayor cuanto menor sea el precio del subyacente, como indica la curva verde de la figura 1, que representa la ganancia del comprador de una opción de venta a 16 con una prima de 2.

Veamos ahora el juego de Epstein. Para jugar sólo hace falta un dado normal, que marcará el precio del subyacente, y unas cuantas cuartillas y bolígrafos. En la versión original, uno de los jugadores debe ejercer el papel de organizador: decide cuántos turnos habrá en el juego, la duración de los mismos, el tipo de contratos y el precio de las opciones. Actúa también como árbitro de las transacciones y como "animador" del mercado cuando éste presenta poca actividad. Si embargo, es también posible prescindir del organizador si los jugadores se ponen de acuerdo al principio de la partida acerca de todas estas cuestiones. Un turno consiste en lanzar el dado y a continuación realizar transacciones entre los jugadores durante el tiempo establecido. Cada jugador tiene una tabla en donde se anotan las transacciones realizadas. Una vez que dos jugadores acuerdan una transacción, anotan en su tabla lo siguiente:

Contrato: el tipo de contrato y el precio de opción o de futuro.

Orden: el número de contratos que se compran o venden. Uno de los jugadores necesariamente tiene que comprar y el otro que vender cada uno de esos contratos.

Prima: es nula en el caso de futuros y es la acordada por los dos jugadores si se trata de opciones.

Estas anotaciones se hacen en las casillas azules de la figura 2. Si un jugador rellena una casilla con un acuerdo de compra, otro tiene que rellenar una casilla con la orden de venta correspondiente. De este modo se realizan las transacciones. No hay límite para el número de contratos y la única forma de "deshacerse" de ellos es realizar una nueva operación de venta, aunque probablemente con una prima o precio de futuro diferente.

Las casillas rojas se rellenan al final del juego, cuando se conoce la puntuación total de todas las tiradas del dado, es decir, el valor del activo subyacente. Si llamamos S a ese valor, las anotaciones se hacen de acuerdo con las siguientes reglas:

Liquidación del contrato: En el caso de futuros es la diferencia entre el subyacente y el precio de futuro; en el caso de la opción de compra, es la diferencia entre S y el precio de opción, si ésta es positiva, y cero si el precio de opción es mayor que S; para opciones de venta es la diferencia entre el precio de opción y S, siempre que esta diferencia sea positiva, y es cero si S es mayor que el precio de opción.

Saldo por contrato: Es lo que ganamos o perdemos en cada contrato, contando con la prima que abonamos al comprarlo o que cobramos al venderlo. Si la orden es de compra, entonces el saldo es la diferencia entre la liquidación y la prima (y coincide para cada contrato con las curvas de la figura 1). Si es de venta, entonces es ese mismo número cambiado de signo, es decir, la prima menos la liquidación.

Saldo total: se obtiene multiplicando el saldo por contrato por el número de contratos intercambiados.

Evidentemente, el ganador es el jugador que obtiene una mayor ganancia total al terminar el juego.

Veamos una partida de cuatro turnos y dos jugadores, Alicia y Bruno, en la que se ha decidido al comienzo que haya futuros y opciones de compra y venta con un único precio de opción de 12. Recordemos que en cada transacción los dos contratantes tienen que acordar el precio de futuro y, en el caso de opciones, la prima. Las transacciones de cada jugador están anotadas en sus respectivas tablas, que se muestran en la figura 2. En el primer turno sale un cinco en el dado. Como quedan otras tres tiradas, el valor esperado del subyacente al final de la partida es 5 + + 10,5 = 15,5 (el valor esperado de una tirada de dado es 3,5). Bruno confía en que las puntuaciones posteriores del dado sean altas y decide ofrecer cinco opciones de venta con una prima de 3 que Alicia acepta. En el segundo turno sale un dos y ahora las previsiones para el subyacente cambian a la baja: el nuevo valor esperado es 14. Para contrarrestar posibles pérdidas si el valor del subyacente baja demasiado, Bruno aprovecha la ocasión para ofrecer futuros a 14 y la

ALICIA

CONTRATO	ORDEN	PRIMA	LIQUIDACION DEL CONTRATO	SALDO POR CONTRATO	SALDO TOTAL
PUT 12	COMPRAR 5	3	0	0 – 3 = -3	-3 × 5 = -15
FUTURO 14	COMPRAR 10	0	17 – 14 = 3	3 – 0 = 3	3 × 10 = 30
CALL 12	VENDER 10	2	17 – 12 = 5	2 – 5 = -3	-3 × 10 = -30
CALL 12	COMPRAR 10	3	17 – 12 = 5	5 – 3 = 2	2 × 10 = 20
			GANANCIA TOTAL		5

BRUNO

CONTRATO	ORDEN	PRIMA	LIQUIDACION DEL CONTRATO	SALDO POR CONTRATO	SALDO TOTAL
PUT 12	VENDER 5	3	0	3 – 0 = 3	3 × 5 = 15
FUTURO 14	VENDER 10	0	17 – 14 = 3	0 – 3 = -3	-3 × 10 = -30
CALL 12	COMPRAR 10	2	17 – 12 = 5	5 – 2 = 3	3 × 10 = 30
CALL 12	VENDER 10	3	17 – 12 = 5	3 – 5 = -2	-2 × 10 = -20
			GANANCIA TOTAL		-5

2. Ejemplo de un juego con cuatro turnos en donde las tiradas han sido 5, 2, 4 y 6. El valor final del subyacente es la suma, es decir, 17

transacción se completa, junto con 10 opciones de compra que Alicia vende con una prima de 2. Finalmente, en el último turno, después de salir un cuatro en el dado, Alicia consigue "deshacerse" de estas 10 opciones de compra, pero esta vez, ya que la puntuación del dado ha sido ligeramente superior a la media, tendrá que pagar una prima mayor.

Un buen jugador ha de tener la habilidad de saber fijar la prima de las opciones. En realidad, es posible calcular la liquidación media de un contrato de opción y con ella conocer la prima más justa. Un programa de ordenador puede hacer esta tarea, calculando las probabilidades de todas las posibles tiradas del dado. Por ejemplo, para una opción de venta a 12 cuando quedan dos tiradas y el subyacente vale en ese momento 7, el "precio justo" sería de 2,28. Una posible variación del juego es contar con este programa y calcular con él las primas. ¿Es posible, incluso utilizando sólo "transacciones justas" diseñar estrategias que venzan al contrincante?

Los precios, primas y fechas de vencimiento de los derivados de verdad se detallan en cualquier periódico. Pero tengan cuidado: en la vida real, calcular el precio de una opción es bastante más complicado. Robert Merton y Myron Scholes ganaron el Nobel de economía en 1997 por elaborar, junto con el fallecido Fischer Black, una teoría para el cálculo de la prima de los productos derivados, la *ecuación de Black-Scholes*. Esta ecuación, junto con variantes que han aparecido posteriormente y siguen apareciendo, es utilizada por los analistas de los mercados financieros. Pero también hay que recordar que algunos de estos analistas, armados de todo este bagaje teórico, han conducido a grandes fondos de inversión a sonoras bancarrotas.

La solución del problema planteado el mes pasado es: el jugador lanza el dado al menos dos veces; si saca CARA-CRUZ es CARA y si saca CRUZ-CARA es CRUZ —ambas opciones tienen la misma probabilidad, *p(1-p)*—. Si saca CARA-CARA o CRUZ-CRUZ, vuelve a empezar.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Juegos equitativos con dados y monedas trucadas

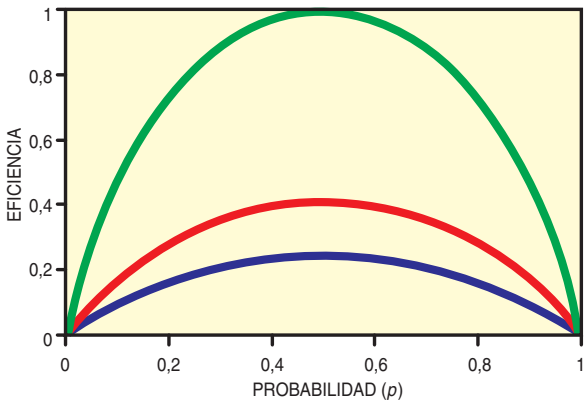
En la sección del pasado mes de abril propuse a los lectores el siguiente problema: ¿es posible diseñar un juego de azar equitativo utilizando una moneda trucada, aun sin conocer las probabilidades de que salga cara o cruz al lanzarla? Von Neumann enunció este problema en 1957; le dio una solución que ya era conocida popularmente y que me han hecho llegar varios lectores. Lanzamos dos veces la moneda: si sale cara-cruz, entonces gana un jugador; si sale cruz-cara, gana el otro; finalmente, si sale cara-cara o cruz-cruz se repite el doble lanzamiento. Las dos posibilidades cara-cruz y cruz-cara son evidentemente simétricas y por tanto tienen la misma probabilidad de ocurrir, se haya trucado como se haya trucado la moneda.

Von Neumann enunció el problema relacionándolo con la generación de *bits* sin sesgo. Recordemos que un *bit* puede tomar sólo dos valores, 0 y 1. Una moneda se puede considerar como un generador de bits si asociamos, por ejemplo, un 0 a cara y un 1 a cruz. Si la moneda está trucada, es decir, si la probabilidad de que salga cara es p, siendo p distinto de 1/2, entonces los bits generados tendrán un cierto sesgo: la fracción de unos en la serie generada por la moneda será aproximadamente igual a p si la lanzamos un número muy grande de veces. El problema de conseguir un juego equitativo a partir de una moneda trucada es entonces equivalente al de conseguir bits sin sesgo a partir de bits sesgados.

La solución expuesta, aunque correcta, es bastante ineficiente. Supongamos que lanzamos la moneda $2N$ veces (recordemos que el número de lanzamientos ha de ser par). ¿Cuántos bits sin sesgo conseguiremos, en media? De cada par de tiradas, obtenemos un bit sin sesgar sólo si el resultado es cara-cruz o cruz-cara. La probabilidad de cada uno de estos resultados es $p(1-p)$ y la probabilidad de que aparezca uno u otro es $2p(1-p)$. Por tanto, el número medio de bits sin sesgo que obtendremos después de $2N$ tiradas será $N_{\text{bits}} = N[2p(1-p)]$. Podemos definir la "eficiencia", E, del método como el número de bits sin sesgar que conseguimos en media por cada lanzamiento de moneda, es decir:

$$E = \frac{N_{\text{bits}}}{2N} = p(1-p).$$

La curva azul de la figura 1 representa esta eficiencia, que es muy baja cuando p se acerca a cero o a uno. La razón es que, cuando la moneda está tan cargada que apenas sale cruz, será muy difícil conseguir bits que no estén sesgados: no se puede obtener incertidumbre a partir de algo que es prácticamente determinista. Parece como si hubiera una "ley de conservación de la incertidumbre", parecida a la ley de la conservación de la energía. De hecho, puede formularse esta ley con precisión matemática, y dicha formulación constituye una de las bases de la teoría de la información de Claude Shannon, a la que hemos dedicado ya algunos artículos (véanse los *Juegos Matemáticos* de agosto y septiembre de 2001). De hecho, más que una ley de conservación de la incertidumbre, se trata de una ley de conservación de la información, ya que información e incertidumbre son equivalentes en la teoría de Shannon.



1. La eficiencia del método de las parejas (en azul) *y del método de los grupos de cuatro* (en rojo). *En verde se puede ver la eficiencia máxima dada por la teoría de la información*

Aunque parezca una perogrullada, una serie de, digamos, 1000 bits sin sesgo contiene exactamente 1000 bits de información; una de 1000 bits sesgados, en cambio, contendrá menos. Un ejemplo límite sería una serie en donde todos los bits fuesen iguales a cero o a uno, cuyo contenido informativo es prácticamente nulo. Shannon, encontró la siguiente fórmula matemática para calcular la información contenida en una serie de N bits sesgados, en la que el uno aparece con probabilidad p y el cero con probabilidad $1-p$:

$$I = -N[p\log_2 p + (1-p)\log_2(1-p)].$$

Según esta fórmula, la información es nula si p es cero o uno, es decir, si en la serie no hay incertidumbre alguna, mientras que es máxima si p es 1/2, caso en que la información es igual al tamaño N de la serie.

¿Cómo se aplica la teoría de la información al problema de la generación de bits sin sesgo con la moneda trucada? Nuestra moneda genera una serie de bits con un contenido informativo dado por la fórmula de Shannon. Por otro lado, una simple manipulación de la serie no puede crear nueva información. Por tanto, con N tiradas de la moneda podremos obtener, como máximo, I bits aleatorios. Tenemos por tanto un valor máximo para la eficiencia:

$$E_{max} = \frac{I}{N} = -[p \log_2 p + (1-p)\log_2(1-p)].$$

Este valor es un límite superior que ningún método, por eficiente que sea, podría superar. Nuestro principio de conservación se parece más ahora a la segunda ley de la termodinámica que a la conservación de la energía: manipulando de modo determinista una serie de datos es imposible aumentar la cantidad de información que contiene.

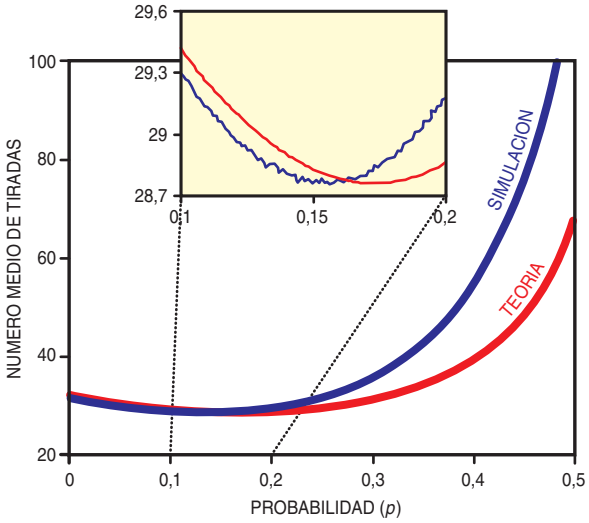
En la figura 1 se puede ver el valor máximo de la eficiencia en función de *p*. La eficiencia del método que hemos expuesto al principio del articulo está muy por debajo de la máxima. Por eso decíamos antes que el método es bastante ineficiente. Esto es más evidente en el caso en que *p* es 1/2. Nuestro método rechaza la mitad de las tiradas y, de las que utiliza, obtiene sólo un bit por cada dos tiradas. De modo que la eficiencia es 1/4. Sin embargo, la moneda genera por sí sola bits sin sesgo, de ahí que la eficiencia máxima sea igual a uno. Claro que la principal virtud del método descrito es que valga para cualquier valor de *p*. ¿Se puede diseñar un método que, siendo independiente de *p*, alcance la eficiencia máxima?

Un método más eficiente —una versión simplificada de una idea que me ha enviado el profesor de la Universidad de Valencia Rafael Pla— utiliza grupos de cuatro tiradas y obtiene con ellos la máxima información posible. El método se resume en la tabla de la derecha, en la que la cara de la moneda se representa por una C y la cruz por una X. Observe que algunos grupos "codifican" dos bits sin sesgo mientras que otros codifican sólo uno. Los grupos en donde aparecen cuatro caras o cuatro cruces se desechan, como en el caso de la generación de bits a partir de parejas. La eficiencia del método está todavía lejos de la máxima, como se puede ver en la figura 1, en donde se lo representa mediante la curva roja. Para alcanzarla, haría falta considerar grupos cada vez más grandes.

BITS SESGADOS	BITS SIN SESGAR	BITS SESGADOS	BITS SIN SESGAR
CCXX	1	CXCC	11
XXCC	0	CCXC	10
CXCX	11	CCCX	01
XCXC	00	CXXX	11
XCCX	01	XCXX	00
CXXC	10	XXCX	01
XCCC	00	XXXC	10

También sobre dados y monedas trucadas, formulaba en la sección del mes de abril la siguiente pregunta: ¿es ventajoso jugar al parchís con un dado trucado en el que el seis tenga más probabilidad de salir que la habitual? La respuesta no es fácil, puesto que el seis en el parchís tiene su parte de cal y su parte de arena: con él avanzamos el mayor número de casillas, pero si sale tres veces consecutivas tenemos que devolver la ficha a casa. En aquel artículo, daba una solución bastante parcial, obtenida mediante un análisis de un parchís muy simplificado. Raquel Alvarez, César Fernández y Armando Relaño, estudiantes de la Universidad Complutense de Madrid, me han enviado una solución más completa. Han simulado un juego de parchís "casi" real. En él hay sólo un jugador que tiene que recorrer, con una única ficha, las 72 casillas que separan la casa de la meta. La ficha avanza según la puntuación de un dado en el que el seis tiene una probabilidad de salir igual a *p* y el resto de los números tienen una probabilidad de salir igual a $(1-p)/5$. En la simulación, se contabiliza el número total de tiradas necesario para llegar a la meta, teniendo en cuenta las siguientes reglas: cuando sale un seis, además de avanzar la ficha, el jugador vuelve a tirar sin que se contabilice la nueva tirada; si sale un seis tres veces seguidas, la ficha vuelve a casa; para salir de casa es necesario sacar un cinco; finalmente, hay que llegar a la meta con la puntuación exacta, y si la puntuación es superior, entonces la ficha "rebota".

Con estas reglas, han jugado un millón de veces para cada valor de *p* y calculado el número medio de tiradas necesario para llegar a la meta. El resultado puede verse en la curva azul de la figura 2. He dibujado también en la misma figura el número medio de tiradas que se obtiene con la teoría, más simple, que expuse en el artículo de abril, en la que una ficha avanza la puntuación del dado y retrocede 36 casillas cuando salen tres seises consecutivos. Las dos curvas difieren bastante para valores altos de *p*, debido a que en ese caso reglas como la del cinco para sacar ficha o la de llegada exacta a la meta cobran una mayor importancia y no se tuvieron en cuenta en el cálculo teórico de abril.



2. Número medio de tiradas para alcanzar la meta en el parchís, en función de la probabilidad p de sacar un seis en el dado. En rojo, la teoría descrita en el artículo de abril de 2001. En azul, la simulación de Raquel Alvarez, César Fernández y Armando Relaño

Lo curioso es que el mínimo de ambas curvas está cerca de 1/6, es decir, del valor correspondiente a un dado no trucado, como se ve en la ampliación de la gráfica entre 0,1 y 0,2. Esto significa que la regla de volver a casa con tres seises prácticamente neutraliza cualquier ventaja que pudiera surgir al cargar el dado hacia el seis. Es como si, a lo largo de la historia del parchís, cuyos orígenes se remontan a la India, la regla hubiera surgido para lograr que el juego fuera equitativo incluso con dados trucados. Aunque esta afirmación no es del todo cierta, porque un dado en el que el cinco

salga con una probabilidad alta sí es claramente ventajoso. Es probable que un parchís más justo sería aquél en el que la ficha saliera de casa cuando se obtuviese un uno en el dado.

JUEGOS MATEMÁTICOS

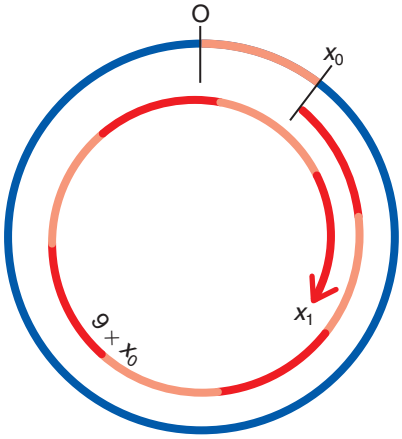
Juan M. R. Parrondo

Caos, determinismo y voluntad

En el problema azar-determinismo parece no haber espacio alguno para eso que llamamos "voluntad". Un suceso, o bien está determinado por el pasado o bien es aleatorio. ¿Cómo surge, entre estas dos posibilidades, la sensación que experimentamos ante un acto voluntario? Cuando decidimos continuar leyendo este artículo o dejar de hacerlo aquí, tenemos la íntima (y legítima) impresión de estar tomando una decisión: el acto no está determinado ni es aleatorio, sino que nace de nuestra propia voluntad. Nos sentimos "creadores" de esa decisión, capaces de cambiarla en el último segundo, lo cual parece incompatible con el determinismo; y somos conscientes en todo momento de ella, de modo que tampoco puede explicarse como un suceso completamente aleatorio. Experimentamos la voluntad como una especie de "generador" insertado en la cadena de la causalidad, cuya naturaleza escapa a la dicotomía azar/determinismo.

Para muchos científicos y pensadores, la teoría del caos determinista ha supuesto una "tercera vía", en la que la voluntad puede tener una mejor cabida. Sin abandonar el determinismo, nos muestra que puede haber una riqueza de comportamiento suficiente como para albergar la sensación subjetiva de voluntad.

Una de las ideas clave de la teoría del caos determinista es la llamada *sensibilidad a las condiciones iniciales*, conocida más popularmente como *efecto mariposa*. El nombre proviene de la imagen que utilizó el meteorólogo E. Lorenz para explicar el comportamiento caótico del clima: el batir de las alas de una mariposa en China puede provocar un huracán en el Caribe. El efecto se puede observar en un sistema determinista muy sencillo. Supongamos que un hombre camina a lo largo de una circunferencia que mide un kilómetro (*véase la figura 1*). Cada día camina, en el sentido de las agujas del reloj, nueve veces la distancia que le separa del punto O, tomada a lo largo de la circunferencia. El comportamiento del hombre está así completamente determinado por su posición inicial. Si x_0 es la distancia (en kilómetros) que le separa



Un recorrido circular que genera un comportamiento caótico

del punto O inicialmente, después de caminar durante el primer día se encontrará a una distancia:

$$x_1 = \text{Parte decimal de } (10 \times x_0)$$

Veamos por qué. En el primer día, recorre $9x_0$. Como ya le separaba de O una distancia x_0, su posición con respecto a O sería $10x_0$ si anduviera en línea recta. Sin embargo, hay que tener en cuenta que se mueve sobre una circunferencia de un kilómetro, de modo que da una vuelta cada vez que recorre un kilómetro. Para cancelar el efecto de estas vueltas, se toma simplemente la parte decimal de $10x_0$.

En los días sucesivos, el comportamiento es el mismo, de modo que, si x_n es la distancia que separa al hombre de O en la noche del día n-ésimo, entonces:

$$x_{n+1} = \text{Parte decimal de } (10 \times x_n)$$

Esta ecuación es bastante más sencilla de lo que parece a simple vista. Recordemos que multiplicar por 10 un número no es más que correr su coma decimal un lugar hacia la derecha. Por otra parte, tomar la parte decimal de un número no es más que hacer cero su parte entera. Por lo tanto, para obtener la posición en el día $n + 1$ lo único que hay que hacer es eliminar la primera cifra decimal de la posición en el día n. Por ejemplo, si la posición inicial es 0,4539..., en la noche del primer día será 0,539..., en la del segundo día 0,39..., y en la del tercer día, 0,9... ¿Cuál es la posición en la noche del cuarto día? Para calcularla, necesitaríamos saber con más precisión la posición inicial de nuestro individuo. Como puede verse, a lo largo de los días, el desarrollo decimal de la posición inicial del individuo se va "desplegando", de modo que cada vez necesitamos conocer cifras más alejadas del desarrollo decimal de x_0. Para calcular la posición al cabo de trece días necesitaríamos conocer la posición inicial con una precisión cercana al tamaño de un átomo, que es del orden de 10^{-13} kilómetros. Esto es evidentemente absurdo: la misma noción de posición deja de tener sentido con una precisión tan grande, puesto que, aunque fijáramos un punto del individuo, su centro de masas, digamos, las fluctuaciones serían mayores que la precisión requerida. Sin embargo, lo interesante de nuestro ejemplo no estriba en si es posible o no realizarlo físicamente, sino en lo que puede ilustrar acerca de la nueva visión del determinismo que ofrece la teoría del caos.

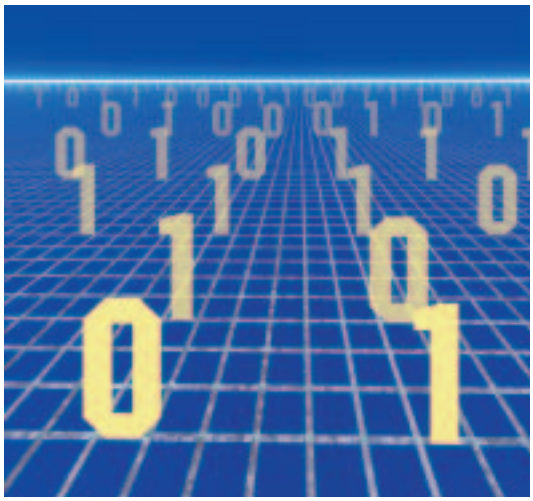
La sensibilidad a las condiciones iniciales resulta evidente en nuestro sistema: si dos hombres parten de posiciones iniciales muy cercanas, que difieren en una cantidad des-

conocida pero del orden de la centésima de milímetro, al cabo de ocho días la situación de uno no tendrá nada que ver con la del otro. El sistema amplifica cualquier desviación inicial, por pequeña que sea, si se espera suficiente tiempo, al igual que la dinámica caótica de la atmósfera es capaz de amplificar perturbaciones minúsculas, como el batir de las alas de una mariposa.

¿Qué tiene que ver el efecto mariposa con la voluntad y el problema que exponíamos al comienzo del artículo? La respuesta es que un sistema caótico puede desplegar a lo largo del tiempo la información infinita contenida en su condición inicial. Para hacernos una idea de lo que esto supone, vamos a recurrir a lo que bien podría llamarse la *Biblioteca de Babel Multimedia*.

La Biblioteca de Babel es un relato de Borges que recoge una vieja idea: la posibilidad de que exista una biblioteca con todos los libros de, pongamos, 300 páginas que pudieran escribirse con las 28 letras del alfabeto más los espacios y signos de puntuación. El número de volúmenes sería inmenso aunque finito. La biblioteca tendría pasillos y estanterías de los cuales sólo podría recorrerse una ínfima parte a lo largo de toda una vida. Aunque la biblioteca contendría todos los libros escritos y por escribir, todos los tratados filosóficos y todas las posibles novelas, la mayoría de los volúmenes estarían repletos de páginas sin sentido. Borges dibuja la biblioteca como un recinto kafkiano, que contiene grandes revelaciones pero prácticamente inaccesibles. Explorar la biblioteca es como explorar un universo que, como el nuestro, contiene más enigmas que respuestas.

La Biblioteca de Babel Multimedia es una versión modernizada, que contiene todos los posibles DVD. Es de nuevo un número inmenso y finito. En ella están todos los libros, y todas las películas, todos los vídeos y todas las piezas musicales interpretadas de todas las formas posibles, lo que hace de ello algo un poco más perturbador que la biblioteca original de Borges. Se encuentra en ella el vídeo de su primera comunión, el vídeo en el que usted está leyendo este artículo y el vídeo en el que comenzó a leerlo pero lo abandonó en el primer párrafo.



El infinito puede desplegar todas las posibilidades

Lo sorprendente es que un número escogido al azar entre 0 y 1 contiene en su desarrollo decimal toda la Biblioteca de Babel Multimedia. El matemático francés Émile Borel introdujo el concepto de *números normales*, números en cuyo desarrollo decimal están todas las posibles subcadenas de dígitos. En otras palabras, si tomamos un número normal, lo expresamos en binario y cortamos fragmentos equivalentes a la extensión de un DVD, obtendremos una réplica (de hecho obtendremos infinitas réplicas) de la Biblioteca de Babel Multimedia. Más aún, Borel demostró que la mayoría de los números entre 0 y 1 son números normales, de modo que un número escogido al azar será normal con probabilidad uno.

Volvamos a nuestro hombre que camina a lo largo de la circunferencia. Si su posición inicial es un número normal, las que irá ocupando al final de cada día desplegarán una información equivalente a la Biblioteca de Babel. Por supuesto, hay que esperar un número inmenso de días para obtener el primer DVD de la Biblioteca, pero lo importante es que toda esa información se halla contenida, *desde el principio*, en la posición inicial del individuo.

Si la mayoría de los sistemas físicos, químicos, biológicos y neuronales son caóticos, estarán a su vez desplegando en su evolución el contenido informativo de unas condiciones iniciales que pueden ser números normales; además, lo estarán haciendo de forma bastante más rápida que el individuo de nuestro ejemplo. Aunque no sepamos exactamente cuál es esa información y cuál es la dinámica que rige dicho despliegue, de lo que no hay duda es de que la imagen que teníamos del determinismo cambia por completo cuando se consideran estas posibilidades. Que lo que voy a hacer mañana esté descrito con todo detalle en algún volumen de la Biblioteca de Babel es una necesidad lógica, que no contradice el hecho de que yo efectivamente decida lo que voy a hacer. Las dinámicas caóticas, aun siendo deterministas, van desplegando la información infinita contenida en las condiciones iniciales, como el bibliotecario de Borges va extrayendo volúmenes de las estanterías. En esa exploración hay bastante más espacio para albergar la sensación subjetiva de voluntad.

No quiero decir con ello que la teoría del caos resuelva el problema. El problema de la voluntad, con el aún más fundamental de la conciencia, son probablemente los más difíciles a los que se enfrenta la ciencia. Estamos lejos de poder dar una explicación científica a la experiencia que uno tiene de sí mismo, de detallar los procesos y mecanismos que hacen que uno pueda tener conciencia de sí mismo, de ser y de decidir. E incluso puede que tal descripción no sea posible dentro de un marco puramente científico. Pero la teoría del caos sí nos está ofreciendo una idea del determinismo mucho más amplia, en la que las cadenas de causalidad contienen una riqueza inesperada y son capaces de desplegar en tiempo finito una inmensa cantidad de información, igual que el huracán del Caribe contiene, en la suma de causas que lo provocan, detalles ínfimos e imperceptibles como el lejano batir de las alas de una mariposa.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Repartir escasez

Alicia y Bruno están de enhorabuena. Un tío lejano les ha declarado herederos únicos en su testamento. En él indica que Alicia debe recibir 120.000 euros y Bruno 40.000. Pero pronto se descubre que su patrimonio asciende sólo a 100.000 euros. El albacea les propone que el reparto se haga de modo proporcional. El deseo del tío es que Alicia reciba el triple que Bruno. Por tanto, lo más correcto es dar 75.000 a Alicia y 25.000 a Bruno. De este modo se respeta escrupulosamente la proporción 3 a 1.

Sin embargo, el astuto abogado de Alicia no está conforme: la pérdida total de los dos herederos, es decir, la diferencia entre lo que esperaban recibir y lo que van a recibir realmente, es de 60.000 euros. Lo más justo sería repartir esta pérdida por igual: que cada uno pierda 30.000 euros con respecto a lo testado. Siguiendo este criterio, Alicia debe recibir 90.000 euros y Bruno 10.000. El albacea está ahora bastante confundido ¿Cuál de las dos propuestas es la más justa?

En matemáticas, a este problema se le conoce como el "de la bancarrota" y, aunque parezca simple, varios grupos de investigación trabajan en la actualidad sobre él. Se denomina así porque aparece también cuando una empresa quiebra y el total de su patrimonio es inferior a su deuda. ¿Cómo repartir el patrimonio entre los distintos acreedores, a modo proporcional a su deuda o de modo que pierdan todos la misma cantidad de dinero?

El problema, en su formulación más general, es el siguiente: se tiene que repartir una cantidad X entre N individuos, cada uno reclama una cantidad c_i y la suma de todas las cantidades reclamadas $C = c_1 + c_2 + \ldots + c_{N-1} + c_N$ es mayor que la cantidad disponible X. Hay que encontrar una regla para repartir X entre los N individuos, es decir, asignar una cantidad x_i a cada individuo i de forma que la suma sea X.

La regla propuesta por el albacea se denomina *regla proporcional* y consiste en asignar:

$$x_i = \lambda c_i \text{ siendo } \lambda = \frac{X}{C} < 1$$

La regla propuesta por el abogado de Alicia se llama *regla de pérdidas iguales* y prescribe la asignación:

$$x_i = \max\{0, c_i - l\}$$

en donde l se calcula de modo que la suma de las asignaciones x_i sea igual a X. Con esta regla, las pérdidas de los reclamantes son todas iguales a l, salvo para los que reclaman una cantidad tan pequeña que diera una asignación negativa al restarle l.

Una tercera regla es la *regla de ganancias iguales*, que reparte el total X por igual entre todos los individuos, siempre que ninguno reciba más de lo que reclama. Matemáticamente se escribe así:

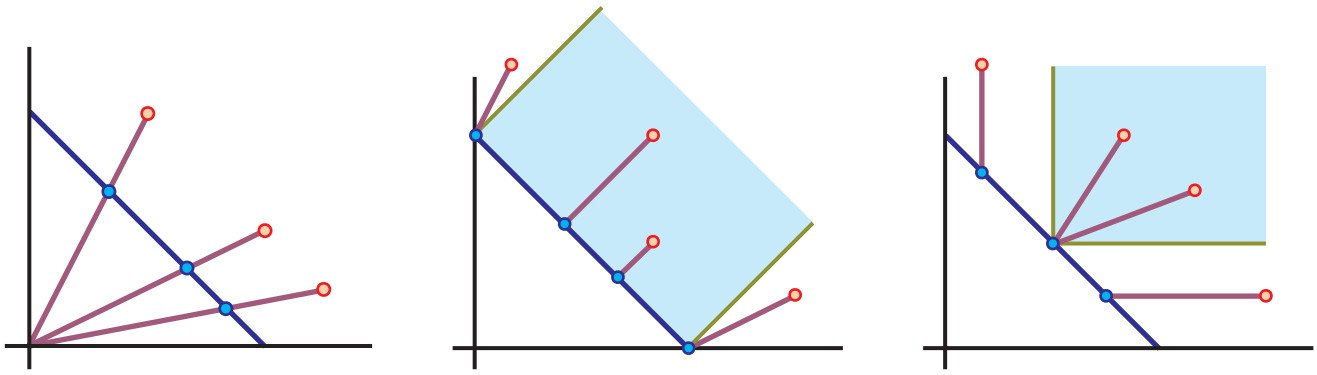
$$x_i = \min\{c_i, r\}$$

en donde r se calcula imponiendo que la suma de las asignaciones x_i sea igual a X.

Cuando se trata de sólo dos individuos, el problema tiene una interpretación geométrica sencilla. El par $c = (c_1, c_2)$ es un punto en el plano, que se dibuja en color naranja en las gráficas de la figura 1; mientras que las posibles soluciones $x = (x_1, x_2)$, tales que $x_1 + x_2 = X$, forman una recta inclinada, que se ha dibujado en azul. Una regla es entonces un modo de asignar a un punto c un punto x (*de color azul en la figura 1*) en la recta de posibles soluciones. Por ello, Antonio Villar, matemático de la Universidad de Alicante que trabaja en el problema de la bancarrota, suele referirse a él como "la historia de un punto y una recta".

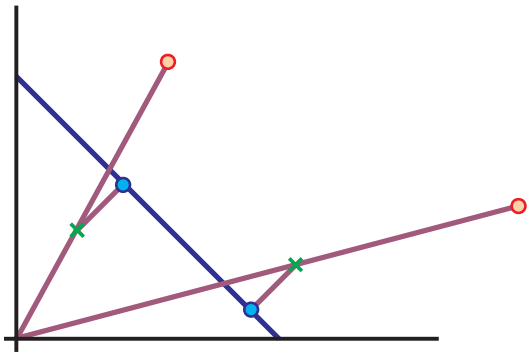
En las tres gráficas de la figura 1 vemos cómo actúan cada una de las tres reglas anteriores. Para la

1. Las tres reglas descritas en el texto. De izquierda a derecha: la regla proporcional, la de pérdidas iguales y la de ganancias iguales. En todas las gráficas los puntos de reclamaciones son los naranjas y los de asignaciones los azules



regla proporcional se traza una recta que une el origen de coordenadas con el punto *c*. El punto *x* es la intersección entre esta recta y la recta de posibles soluciones. En la regla de pérdidas iguales, *x* es el punto de la recta de posibles soluciones más cercano a *c*, lo cual parecería indicar que esta regla es incluso más justa que la proporcional. Finalmente, la regla de ganancias iguales elige el punto medio de la recta de posibles soluciones siempre que *c* esté en la región sombreada. Cuando el *c* está fuera de esa región, la regla asigna el punto de la recta de soluciones que está en la misma vertical u horizontal de *c*, tal y como muestra la figura 1.

Cada una de estas reglas tiene una larga historia. La proporcional fue descrita por Aristóteles, mientras que las de pérdidas y ganancias iguales se encuentran en los códigos de Maimónides del siglo XII. Es fácil comprobar que la regla de pérdidas iguales favorece a los grandes acreedores porque excluye del reparto a los individuos que reclaman cantidades pequeñas. Por el contrario, la de ganancias iguales favorece a los acreedores pequeños, ya que éstos consiguen el total de lo reclamado. Esta regla se aplica en situaciones reales, como la bancarrota de intermediarios financieros, en donde se saldan primero las deudas contraídas con los clientes pequeños por dos razones: porque suelen ser los menos pudientes y a los que más afecta no cobrar la deuda y, en segundo lugar, porque no pueden considerarse responsables de la bancarrota, ya que si los *N* deudores reclamaran cantidades inferiores a *X*/*N* ésta no se habría producido.



2. Representación de la regla del Talmud. La cruz verde es el punto medio entre el origen y el punto c *de reclamaciones. El punto* x *de asignación es el más cercano de la recta azul a la cruz verde*

A estas tres reglas se añade una cuarta un poco más complicada. La regla surgió de considerar el problema de la bancarrota como un problema de teoría de juegos, pero se ha demostrado posteriormente que coincide con las prescripciones que da el Talmud para distintos casos particulares. De modo que la regla es tan antigua como las tres anteriores. La regla del Talmud (también llamada *nucleolo*) es una sutil combinación de la de pérdidas iguales y ganancias iguales. Consiste en dividir por la mitad el total de reclamaciones. Si *C*/2 es mayor que el capital a repartir *X*, entonces se aplica la regla de pérdidas iguales pero considerando sólo la mitad de la reclamación de cada individuo. Si *C*/2 es mayor que *X*, entonces se reparte *X* satisfaciendo la mitad de la reclamación de cada individuo y el resto se reparte según la regla de ganancias iguales. Matemáticamente se escribe así:

$$x_i = \begin{cases} \min\{c_i/2, r\} \text{ si } X \le C/2 \\ c_i/2 + \max\{0, c_i/2 - l\} \text{ si } X \ge C/2 \end{cases}$$

Los valores de *r* y *l* se escogen, en cada caso y al igual que en las reglas anteriores, de modo que la suma de asignaciones sea igual a *X*. La representación gráfica de la regla del Talmud se muestra para un par de ejemplos en la figura 2: se traza una línea entre el origen y *c* y se marca el punto medio de la misma; a partir de este punto, se busca el más cercano de la recta de soluciones posibles, siempre que las asignaciones sean positivas y menores que las reclamaciones. Como afirman Carmen Herrero y Antonio Villar, la regla del Talmud parece basarse en el principio psicológico "más de la mitad es prácticamente el todo, mientras que menos de la mitad es prácticamente nada". Así, la regla trata de satisfacer la mitad de las reclamaciones, al menos para los acreedores pequeños.

En la herencia de Alicia y Bruno, la regla del Talmud se aplicaría de la siguiente forma. Las reclamaciones de los dos herederos son 120.000 y 40.000 euros, respectivamente. La mitad de la reclamación total es por tanto 80.000 euros, menos que el capital a repartir, *X* = 100.000 euros. Por tanto, repartimos 80.000 euros, dando 60.000 a Alicia y 20.000 a Bruno. Los 20.000 restantes los repartimos mediante la regla de ganancias iguales entre los dos herederos, que ahora reclaman 60.000 y 20.000 euros, respectivamente. Damos, por lo tanto, 10.000 euros a cada uno. En total, Alicia habrá recibido 70.000 euros y Bruno 30.000. Esta solución difiere de las anteriores. En nuestro problema de la herencia, cada regla establece asignaciones diferentes: la proporcional da 75.000 euros a Alicia y 25.000 a Bruno, la de pérdidas iguales 90.000 y 10.000, la de ganancias iguales 60.000 y 40.000, mientras que el reparto con la regla del Talmud sería de 70.000 y 30.000. ¿Cuál es la correcta? No hay una respuesta definitiva. El trabajo del grupo de Antonio Villar y de otros matemáticos trata de encontrar las propiedades básicas de cada una de las reglas que hemos descrito. Lo importante es conocer la variedad de reglas y qué propiedades las caracterizan. De este modo podremos saber cuál es la más adecuada en cada caso. Un ejemplo de indudable importancia práctica es el trabajo de los economistas de la Universidad del País Vasco Carmen Gallastegui, José Manuel Chamorro, Javier Fernández Macho y Elena Iñara, en el que consideran la asignación de cuotas de pesca a los países de la Unión Europea como un problema de bancarrota. En este caso, las reclamaciones serían las cantidades c_i pescadas por cada país en períodos en donde no había restricciones. El total *C* de estas reclamaciones excede a la cantidad *X* de pesca máxima que garantiza la sostenibilidad de los recursos pesqueros. El problema es entonces idéntico al de nuestra conflictiva herencia entre Alicia y Bruno y al de muchas otras situaciones en donde haya que repartir algún bien escaso.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Paradojas y atascos de tráfico

En una ciudad, el barrio residencial *A* está conectado con el barrio de oficinas *B* por dos carreteras, como se muestra en la figura 1. Cada mañana, mil coches se dirigen de *A* a *B;* cada conductor elige uno de los dos posibles caminos tratando de llegar a su destino en el menor tiempo posible.

Los coches tardan en recorrer los tramos *AD* y *CB* 15 minutos. En los tramos *AC* y *DB* hay puentes en los que la carretera se estrecha. Como consecuencia, el tráfico es más lento cuantos más coches atraviesan los puentes. Supongamos que los coches tardan $F_{AC}/100$ minutos en recorrer el trayecto *AC*, siendo F_{AC} el número de coches que esa mañana atraviesan el puente. De la misma forma, se tarda $F_{BD}/100$ minutos en recorrer el trayecto *BD*, siendo F_{BD} el número de coches que esa mañana atraviesan el puente situado entre *B* y *D*. Por ejemplo, si todos los coches deciden ir de *A* a *B* por el trayecto *ACB*, tardarán 1000/100 = 10 minutos en ir de *A* a *B*, más 15 minutos en ir de *C* a *B*. El viaje completo durará 25 minutos. Por supuesto, los conductores avispados al día siguiente elegirán el trayecto *ADB*. Si, por ejemplo, sólo 100 toman ese trayecto, tardarán 15 + 100/100 = 16 minutos en llegar a su trabajo, mientras que los que insisten en el trayecto *ACB* tardarán 15 + 900/100 = 21 minutos.

Está claro que si uno de los dos trayectos es más rápido que el otro, los conductores tenderán a elegirlo en días posteriores. Pero con ello harán que sea más lento. Se llegará a un equilibrio si los tiempos en cada trayecto son iguales, de modo que para nadie suponga una ventaja cambiar su trayecto cotidiano. Este equilibrio coincide con el llamado *equilibrio de Nash* cuando el problema de escoger el trayecto se interpreta como un problema de teoría de juegos: no hay cambio de estrategia individual que permita que jugador alguno aumente su ganancia.

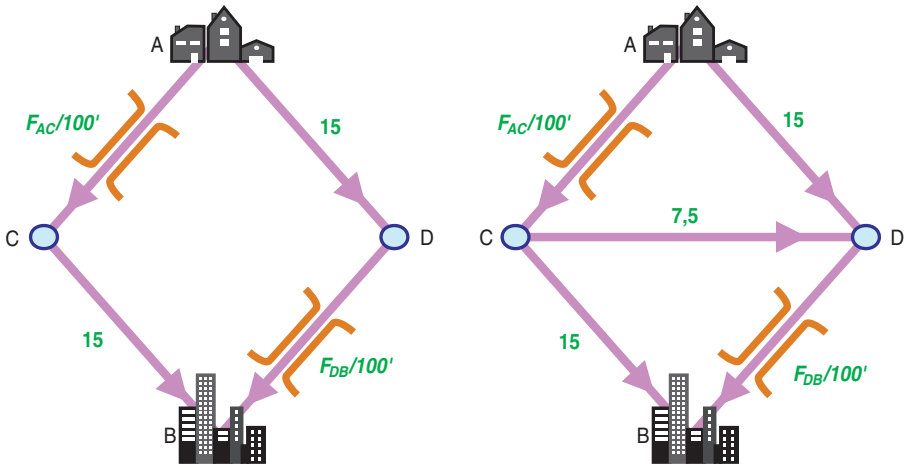
En el caso de nuestra red viaria, el equilibrio se consigue si:

$$T_{ACB} = T_{ABD} \Rightarrow \frac{F_{AC}}{100} + 15 = 15 + \frac{F_{DB}}{100}$$

lo que implica inmediatamente que el mismo número de coches atraviesa los dos puentes cada mañana. Por tanto, en el estado de equilibrio 500 conductores elegirán el trayecto *ACB* y otros 500 el trayecto *ADB*, y el tiempo que invertirán todos ellos en su viaje al trabajo será:

$$T_{ACB} = T_{ABD} = \frac{500}{100} + 15 = 20 \text{ minutos}$$

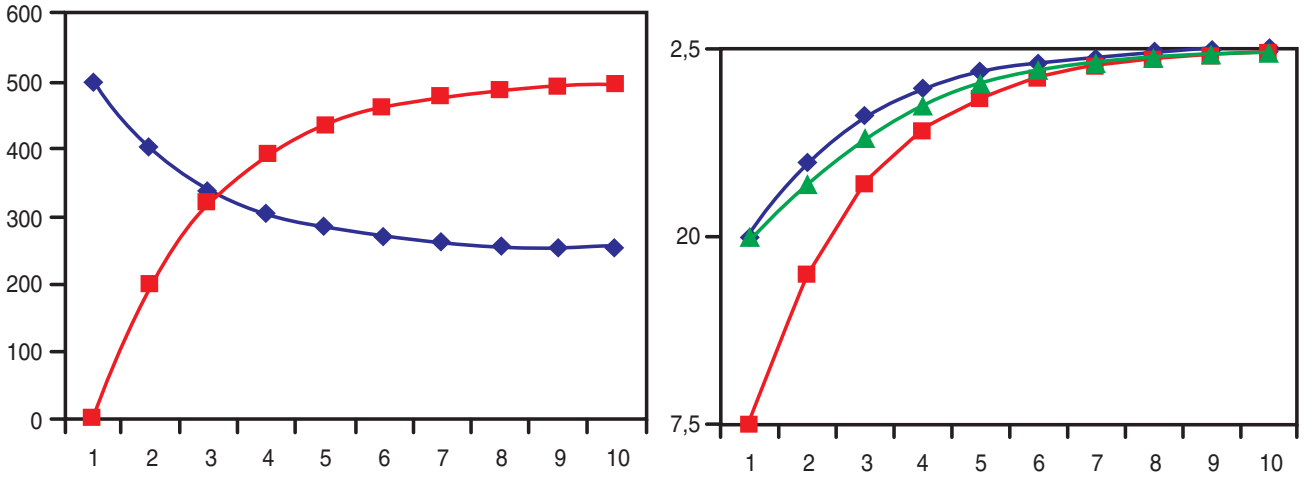
Unos meses antes de las elecciones, el alcalde decide invertir algo del presupuesto municipal en mejorar la red viaria; se proyecta una carretera de un único sentido que una los puntos *C* y *D*. El tiempo medio que se invierte para ir de *C* a *D* por la flamante vía es de 7,5 minutos. La carretera se inaugura en un solemne acto. Al día siguiente, los conductores que viven en el barrio *A* están felices. Los más avispados toman el trayecto *ACDB* y se ahorran más de dos minutos en su viaje diario al trabajo. Sin embargo, al cabo de una semana, el alcalde comienza a recibir quejas de los vecinos: protestan porque ahora tardan más que antes en llegar a *B*. "¿Cómo es posible —se pregunta el alcalde— que haya aumentado el tiempo de *todos* los conductores al abrirse la nueva vía? Al fin y al cabo, lo único que hemos hecho es ofrecer un nuevo trayecto posible. Si la nueva carretera produce más atascos o es más lenta, ¿por qué no vuelven los conductores a utilizar los otros trayectos, como si el tramo *CD* no existiera?"



1. Red viaria entre el barrio residencial A y el barrio de oficinas B. En verde se indican los tiempos necesarios para recorrer cada tramo de la red. La red de la derecha, en donde se ha añadido el tramo CD, resulta ser más lenta que la red de la izquierda.

Nuestro confundido alcalde está siendo víctima de la llamada *paradoja de Braess*, enunciada en 1968 por el matemático alemán Dietrich Braess. Se cree que puede ocurrir en problemas de tráfico, de redes de comunicaciones, circuitos eléctricos e hidráulicos o teoría de juegos. La formulación de la paradoja es muy simple: en ciertas ocasiones, añadir a una red de comunicaciones un nuevo tramo puede disminuir la eficiencia de la red.

Veamos qué ocurre en nuestro ejemplo. Con el tramo *CD* hay tres posibles caminos para ir de *A* a *B*: *ACB*, *ADB* y *ACDB*. El tiempo invertido en cada uno de ellos será:



2. A la izquierda, en azul, se representa el número de coches que optan por el tramo ACB cada día, que son los mismos que optan por el ADB; la curva roja es el número de coches que escogen ACDB. A la derecha: la curva azul es el tiempo que tardan los que han elegido ACD o ADB, la curva roja el tiempo de los que van por ACDB y la curva verde el tiempo medio de todos los conductores

$$T_{ACB} = \frac{F_{AC}}{100} + 15$$

$$T_{ADB} = \frac{F_{DB}}{100} + 15$$

$$T_{ACDB} = \frac{F_{AC}}{100} + \frac{F_{DB}}{100} + 7{,}5$$

La situación de equilibrio se alcanza cuando los tres tiempos son iguales. Al igualar T_{ACB} y T_{ADB}, obtenemos que el número de coches que atraviesa cada puente es el mismo: $F_{AC} = F_{DB}$. Finalmente, igualando T_{ACB} y T_{ACDB}, obtenemos el número de coches que pasa por el puente *DB*:

$$\frac{F_{DB}}{100} + 15 = 2 \times \frac{F_{DB}}{100} + 7{,}5 \Rightarrow F_{DB} = 750 \text{ coches}$$

Por lo tanto, el tiempo medio que tarda cada conductor será:

$$T_{ACB} = T_{ADB} = T_{ACDB} = \frac{750}{100} + 15 = 22{,}5 \text{ minutos}$$

que es mayor que el tiempo medio que tardaban los conductores antes de que se abriera el tramo *CD*. ¿Por qué no pueden, como piensa el alcalde, ignorar el nuevo tramo y utilizar sólo los tramos *ACB* y *ADB*, ya que con ello todo el mundo saldría beneficiado? Para analizar este problema, he hecho una simulación en donde los conductores tienden a cambiar de los trayectos más lentos a los más rápidos. En concreto, he supuesto que cada día se pasa de un trayecto a otro un número de conductores proporcional a la diferencia de tiempos entre los dos trayectos. En la figura 2 se puede ver el comportamiento de los conductores y el tiempo medio que emplean en los primeros días tras inaugurarse la carretera. El primer día nadie toma el trayecto *ACDB*, pero después un número creciente de conductores opta por él abandonando los trayectos *ACB* y *ADB*. Pronto se alcanza la situación de equilibrio en la que los tiempos medios en cada trayecto son iguales y ningún conductor cambia ya de camino.

En la 2 se ilustra también la evolución de los tiempos medios de cada trayecto (*curvas roja y azul*) y el tiempo medio de todos los conductores (*curva verde*). El primer día, puesto que nadie lo elige, el trayecto *ACDB* es bastante más rápido. Por ello, en el segundo día bastantes conductores se pasan a él, congestionando los puentes y aumentando el tiempo de todos los demás. Cuando nos encontramos en la situación de equilibrio, nadie quiere cambiar. Si, por ejemplo, 100 conductores decidieran abandonar el trayecto *ACDB* y repartirse por igual entre los otros dos, cruzarían cada puente 700 coches y los tiempos serían: *TACB* = *TADB* = 22 minutos y *TACDB* = 21,5 minutos. Aunque todos han ganado con este traspaso, los más egoístas tomarían de nuevo el trayecto *ACDB* y se volvería a la situación de equilibrio.

La paradoja de Braess nos indica que la maximización del beneficio individual no conduce necesariamente a la mejor solución. Es por tanto una seria advertencia para los fanáticos de la “mano invisible”, que confían en la eficacia del mercado libre. Como muestra la paradoja de Braess, en ocasiones la solución óptima no surge de la maximización de intereses individuales, sino de medidas globales, como eliminar el tramo *CD* en nuestro sencillo ejemplo.

Pero, ¿ocurre en la realidad la paradoja de Braess? Aunque no hay pruebas contundentes, sí ha habido casos en que se ha detectado un aumento en la eficiencia de una red al cortar alguno de sus tramos. Los más conocidos han tenido lugar en barrios de Nueva York y Stuttgart. Cohen y Horowitz han formulado situaciones físicas inspiradas en la paradoja de Braess: un peso cuelga de una red de muelles y cuerdas suspendida del techo. Cuando se corta una de las cuerdas el peso asciende, contradiciendo la intuición. Han diseñado también un circuito eléctrico cuya resistencia disminuye al cortar una de las conexiones entre los elementos del circuito. Todo ello hace que la paradoja sea un campo de estudio en activo y un fenómeno a tener en cuenta en el diseño de redes de comunicaciones.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Monedas, balanzas e información

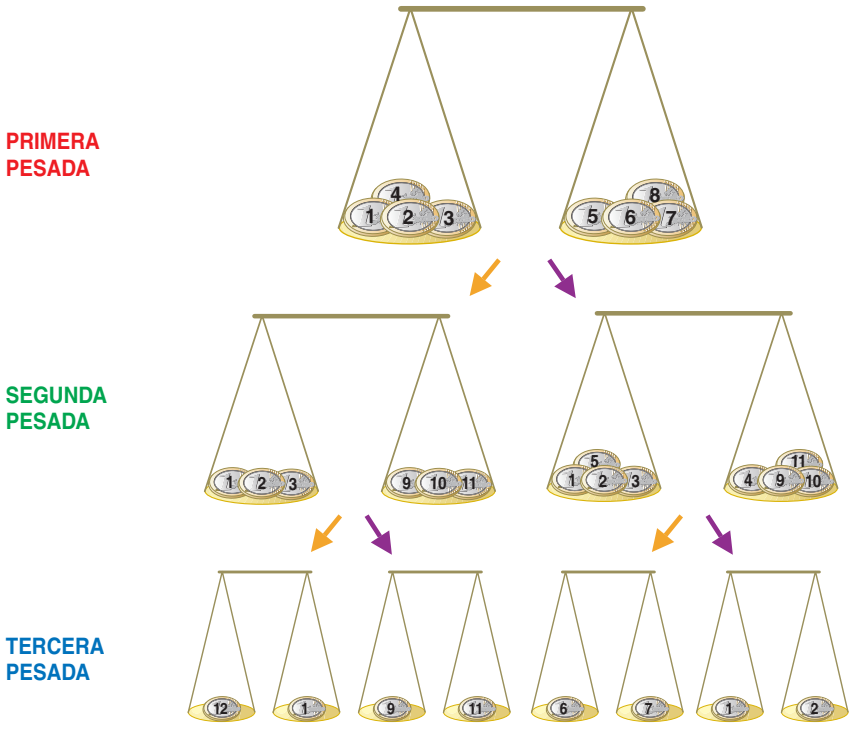
Este mes vamos a analizar un viejo problema de ingenio que los lectores probablemente conocerán. Sin embargo, abordaremos su solución desde un ángulo nuevo, que además nos permitirá ilustrar algunos conceptos de la teoría de la información. El problema consiste en encontrar una moneda falsa entre doce mediante una balanza de dos platillos. La moneda falsa tiene un peso diferente al de las otras, pero no sabemos si es más o menos pesada que éstas. Con estas premisas, debemos encontrar la moneda falsa en sólo tres pesadas.

La solución es bastante complicada y normalmente se llega a ella por prueba y error, o reduciendo el problema a otro más sencillo. Sin embargo, si uno se deja guiar por los principios de la teoría de la información, la solución se construye de forma bastante natural y, sobre todo, sistemática.

Cada pesada tiene tres posibles resultados: que los dos platillos se equilibren, que el derecho pese más que el izquierdo o que sea el izquierdo el más pesado, y cada uno de estos resultados nos proporciona una cierta información sobre las monedas; maximizarla para cualquiera de los tres resultados posibles será la mejor forma de diseñar una pesada. La teoría de la información nos dice que esto se consigue si los tres resultados posibles de la pesada son equiprobables, es decir, si se dan con una probabilidad de 1/3. Este es el principio que guiará en todo momento nuestro diseño de las pesadas: la probabilidad de cada resultado debe ser lo más cercana a 1/3 que podamos.

Supongamos que en la primera pesada colocamos un cierto número de monedas en cada platillo (el mismo número en cada uno) y dejamos algunas fuera. El equilibrio entre los platillos se dará sólo si la falsa se encuentra entre las monedas que dejamos fuera de la pesada. Para que esto ocurra con una probabilidad 1/3, debemos dejar fuera cuatro monedas. Por tanto, en la primera pesada debemos colocar, tal y como se muestra en la figura 1, cuatro monedas en cada platillo, digamos la 1, 2, 3 y 4 en el platillo de la derecha, y la 5, 6, 7 y 8 en el de la izquierda, quedando fuera de la pesada el resto, es decir, la 9, 10, 11 y 12. Con ello aseguramos que los tres resultados posibles de la pesada son equiprobables y que ésta nos proporciona la mayor información posible.

Para entender mejor lo que queremos decir con "la mayor información posible", conviene analizar el problema investigando cómo cada pesada elimina algunas de las diferentes posibilidades que pueden darse en cada situación. En principio, la moneda falsa puede ser cualquiera de las doce, y puede pesar más o menos que el resto. Por lo tanto, tenemos 24 posibilidades. La primera pesada, tal y como la hemos diseñado siguiendo el principio de máxima información, elimina siempre 16 y deja como posibles sólo 8 de ellas. En el caso en que resulte un equilibrio entre platillos, la moneda falsa debe ser la 9, 10, 11 o la 12 y aún no sabemos si es más o menos pesada que el resto. Si la balanza se inclina hacia la derecha,



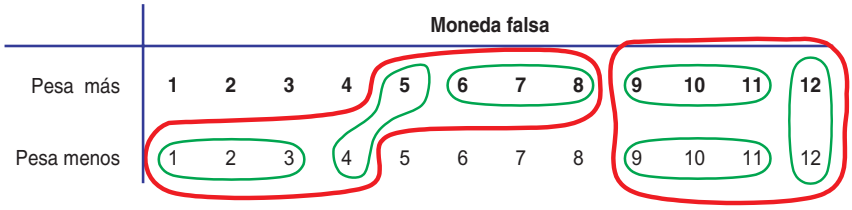
1. Solución del problema de la moneda falsa. La flecha naranja indica la pesada a realizar si el resultado ha sido el equilibrio entre los platillos, mientras que la flecha morada se aplica si la balanza se ha inclinado hacia la derecha. El caso en que la balanza se inclina a la izquierda se puede deducir con facilidad de la tabla

entonces sabremos que la moneda falsa es una de las ocho primeras y que si es la 1, 2, 3 o 4, entonces será menos pesada que el resto, mientras que si es la 5, 6, 7 u 8, tendrá exceso de peso. En todos los casos, las posibilidades compatibles con cada resultado son siempre ocho. En la figura 2 se representan las 24 posibilidades iniciales y, rodeadas por líneas rojas, las ocho posibilidades compatibles con el resultado de equilibrio y con el de desequilibrio hacia la derecha.

Observe que ésta es la máxima reducción de posibilidades que se puede alcanzar con una pesada. Si uno de los resultados eliminara, por ejemplo, 17 posibilidades quedándose sólo con 7, entonces los otros dos serían compatibles con esas 17 posibilidades. Por lo tanto, al menos uno de los resultados sería compatible con 9 o más posibilidades.

La mejor estrategia para el diseño de pesadas es claramente la que consigue la mayor reducción de posibilidades, *para cualquiera de los tres resultados de la pesada*. Lo que acabamos de ver es que esa reducción máxima se consigue cuando los tres resultados son compatibles con el mismo número de posibilidades. Como el número total de posibilidades es 24, cada resultado, en la estrategia óptima, será compatible con un tercio de 24, que es 8. Pero, puesto que la probabilidad de que se dé un resultado no es más que la fracción de posibilidades compatibles con el mismo, vemos que la reducción máxima se consigue siguiendo el principio dictado por la teoría de la información, es decir, haciendo que esas probabilidades sean iguales para los tres resultados.

Para la segunda pesada operamos de la misma forma. El problema es que ahora no es posible conseguir pesadas con resultados equiprobables, ya que el número de posibilidades que quedan es 8, que no es divisible entre 3. Nos podemos acercar si diseñamos pesadas en donde uno de los resultados sea compatible con dos posibilidades, mientras que los otros dos lo sean con 3 posibilidades. En la figura 1 se muestra cómo conseguirlo para los dos casos que pueden darse después de la primera pesada. Finalmente,



2. Reducción de posibilidades en cada pesada. En un principio hay 24 posibilidades. Las áreas rodeadas por las líneas rojas indican las posibilidades que quedan después de la primera pesada en el caso de equilibrio o de desequilibrio hacia la derecha. Las áreas rodeadas por las líneas verdes indican las posibilidades que quedan después de la segunda pesada

cuando sólo quedan dos o tres posibilidades, se puede diseñar una pesada que discierna entre ellas.

En la figura 1 se muestra la solución del problema para cuando los resultados de las pesadas son bien el equilibrio, bien el desequilibrio hacia la derecha. La figura 2 no es más que un pequeño esquema de cómo las pesadas van eliminando posibilidades. Las líneas rojas rodean las que quedan después de la primera pesada. No he incluido la reducción a que conduce el desequilibrio hacia la izquierda para no complicar en exceso la figura. De forma similar, la segunda pesada, tal y como está diseñada en la figura 1, reduciría las posibilidades a los conjuntos rodeados por las líneas verdes. Finalmente, la tercera pesada debe diseñarse de modo que sea compatible con una única posibilidad, obteniendo así la respuesta al enigma.

Lo reseñable en este viejo problema es su relación con la teoría de la información. Los seguidores de la sección de *Juegos matemáticos* recordarán que ya utilizamos esta teoría para explicar por qué en los juegos de preguntas sí/no conviene realizar preguntas en donde las dos respuestas sean equiprobables y la relación entre este tipo de juegos y la compresión de datos en informática (agosto y septiembre de 2001). Cada pesada en el problema que nos ocupa hoy o cada pregunta en un juego proporcionan una cierta cantidad de información que es igual a la disminución de la incertidumbre. Aunque parece que información e incertidumbre son conceptos subjetivos y difíciles de formalizar matemáticamente, la teoría de la información da una definición precisa y cuantitativa de la incertidumbre asociada a una cierta situación, y, en consecuencia, de la información que proporciona una pregunta o una medición acerca de dicha situación. La incertidumbre de una situación dada deberá estar íntimamente relacionada con el número de posibilidades compatibles con ella. La definición precisa de la que hablábamos nos dice que es proporcional a su logaritmo. En el caso de las monedas y las balanzas, la reducción del número de posibilidades compatibles que tiene lugar tras cada pesada, no es más que una reducción de la incertidumbre asociada a las monedas. La cuantificación precisa de esa incertidumbre es lo que nos permite asegurar que las pesadas que proporcionan mayor información son aquéllas en las que los resultados son equiprobables, que es el principio con el que hemos podido encontrar la solución al problema.

Para terminar, invito a los lectores a generalizar el problema. Si disponemos de n monedas, una de las cuales es falsa, ¿cuál será el mínimo número de pesadas que harán falta para descubrirla? No se trata de una pregunta sencilla y no creo que pueda contestarse con absoluta generalidad. Al menos sí es posible encontrar cotas para ese mínimo número de pesadas. En cualquier caso, tratar de contestar a esta pregunta les servirá para profundizar acerca de las relaciones entre incertidumbre e información.

Juan M. R. Parrondo
parr-km0@zenon.fis.ucm.es

JUEGOS MATEMÁTICOS

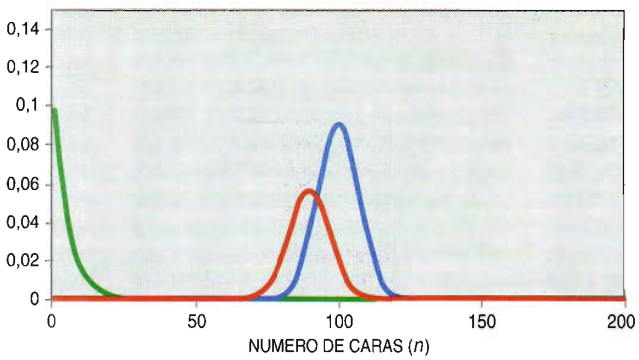
Juan M. R. Parrondo

Ruletas, monedas y entropía

Si en la ruleta de un casino ha salido cinco veces seguidas el mismo número, incluso al jugador más racional le temblaría la mano antes de apostar una vez más a dicho número. Sin embargo, basta un momento de reflexión para darse cuenta de que la bola saltarina de la ruleta, si el crupier es honrado, no recuerda en absoluto los números sobre los que se ha posado en las últimas cinco tiradas. Por tanto, todos los números tienen la misma probabilidad de salir en la sexta.

A pesar de ello, es verdad que la repetición del mismo número seis veces seguidas es algo muy improbable. ¿Significa esto que la bola tiene una cierta memoria de lo ocurrido y evita los números que han aparecido con anterioridad? En absoluto. La clave está en que la probabilidad de que salga el 15 seis veces seguidas es exactamente igual a la de que salga cualquier otra secuencia, por ejemplo, la 8, 22, 13, 4, 9, 17. Cuando decimos que la repetición del mismo número seis veces seguidas es muy improbable, lo que en realidad queremos decir es que es mucho menos probable que salga una secuencia de ese tipo que una secuencia en donde todos los números sean distintos. Aunque cada secuencia tiene la misma probabilidad, hay muchas más secuencias con todos los números distintos que con todos los números iguales. En concreto, hay sólo 37 secuencias con los seis números iguales (recordemos que la ruleta tiene 37 números), mientras que hay $37 \times 36 \times 35 \times 34 \times 33 \times 32 = 1.673.844.480$ secuencias con los números distintos. Por ello, la probabilidad de que salgan seis números iguales es, aproximadamente, de una entre 100 millones, mientras que la probabilidad de que salga una secuencia con todos los números diferentes es del 65 %.

En este ejemplo se muestra claramente la diferencia entre las probabilidades de secuencias concretas y las de *tipos* de secuencias. La probabilidad de cada secuencia concreta es siempre la misma, mientras que las probabilidades de tipos de secuencias pueden ser muy diferentes. En nuestro ejemplo, el tipo de secuencia salir-seis-veces-el-mismo-número es 45 millones de veces más improbable que el tipo de secuencia salir-seis-números-distintos. La probabilidad de un tipo de secuencia se calcula de forma muy sencilla. Basta multiplicar la probabilidad de una secuencia concreta por el número de secuencias de dicho tipo.



1. La curva verde representa la probabilidad de una secuencia concreta con n *caras; la azul, el número de secuencias con* n *caras; la roja, la probabilidad de que salga una secuencia cualquiera con* n *caras. La escala vertical corresponde sólo a la curva roja*

Esta distinción entre secuencias concretas y tipos de secuencia, y el modo como se calculan las probabilidades de cada tipo, es la base de uno de los conceptos más profundos y relevantes de la física moderna: la entropía.

Veamos este cálculo para un caso más sencillo. Supongamos que lanzamos una moneda 200 veces. La moneda está trucada, de modo que la probabilidad de que salga cara es 0,45 y de que salga cruz es 0,55. ¿Cuál es la probabilidad de obtener n caras y 200-n cruces? En este caso, las secuencias concretas no son igualmente probables, puesto que es más probable la cruz que la cara. La probabilidad de sacar, por ejemplo, dos caras en dos tiradas, es $0,45 \times 0,45$. Como las probabilidades se multiplican, la probabilidad de una secuencia concreta que contenga n caras y 200-n cruces será: $0,45^n \times \times 0,55^{200-n}$. El resultado de esta fórmula es la curva verde de la figura 1 (dibujada con un cambio de escala por conveniencia). Como vemos en la figura, la secuencia concreta más probable es la de cero caras y 200 cruces. Aunque parezca increíble, si tuviéramos que apostar por una secuencia concreta entre las $2^{200} = 1,6 \times 10^{60}$ posibles secuencias, lo más inteligente sería apostar a la de 200 cruces seguidas.

Sin embargo, todo cambia si consideramos tipos de secuencias. Secuencias con 200 cruces sólo hay una. Tenemos, sin embargo, 200 secuencias con 1 cara y 199 cruces. Con dos caras y 198 cruces, hay ya 19900 posibilidades y con 100 caras y 100 cruces hay casi 10^{59} posibilidades, un número extraordinariamente grande. En la figura 1, también con una escala adecuada para que pueda verse con claridad, he dibujado en azul el número de posibilidades para cada tipo de secuencia.

Finalmente, la probabilidad de cada tipo es el producto de la probabili-

dad de la secuencia y del número de posibilidades, y el resultado lo he dibujado en rojo en la misma figura. La curva roja tiene su máximo en $n = 90$. Este resultado era esperable: puesto que la probabilidad de que salga cara es 0,45, en un número grande de tiradas deberán salir un 45 % de caras y un 55 % de cruces, y el 45 % de 200 es 90.

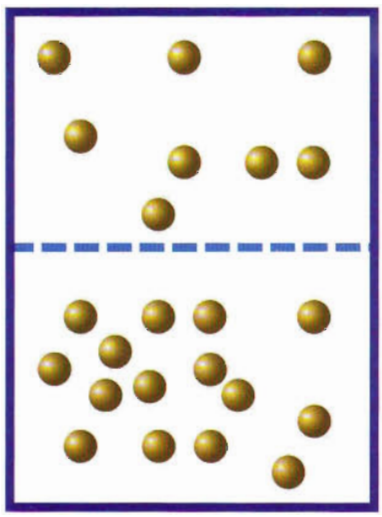
Lo interesante es que las secuencias concretas que constan de 90 caras y 110 cruces son mucho menos probables que la secuencia con 200 cruces, como se ve en la curva verde. La probabilidad de cada tipo, es decir, la curva roja, es el resultado de la combinación de las otras dos curvas. A finales del siglo XIX, el físico austríaco Ludwig Boltzmann utilizó esta combinación para resolver uno de los enigmas más profundos de la física: conciliar el comportamiento del mundo macroscópico con las leyes que rigen el mundo microscópico de los átomos y las moléculas. Cada uno de estos mundos se comporta de manera diferente: en el microscópico el movimiento no cesa, no hay fricción, la energía no se pierde, no hay una flecha del tiempo que distinga el pasado del futuro; en el macroscópico, los cuerpos sufren fuerzas de fricción y tienden a pararse, la energía se disipa en forma de calor inutilizable y hay infinidad de procesos irreversibles, como cuando se hace añicos una copa de cristal al caer al suelo.

Boltzmann se dio cuenta de que la diferencia entre el mundo microscópico y el macroscópico era en realidad una diferencia "de mirada", y que esa diferencia, aun tan simple y sutil, podía explicar los distintos comportamientos de cada uno de los mundos. Cuando miramos el mundo microscópico vemos el movimiento detallado de cada partícula que nos rodea, mientras que cuando miramos el macroscópico sólo vemos los comportamientos colectivos.

Estos dos tipos de "mirada" se dan también en nuestro ejemplo de las secuencias. Un jugador "microscópico" apostará a qué secuencia concreta saldrá en 200 tiradas de una moneda. Por el contrario, uno "macroscópico" será el que apueste al número total de caras de la serie; sólo le interesa el *tipo* de secuencia.

En un sistema físico, llamamos estados microscópicos a las secuencias concretas y estados macroscópicos a los tipos de secuencias. La probabilidad de que se dé un estado microscópico es mayor cuanto menor es su energía. Pero esa probabilidad es la de un estado microscópico concreto, igual que la curva verde de la figura 1 es la probabilidad de una secuencia concreta. Al pasar al mundo macroscópico las cosas cambian radicalmente, igual que en el ejemplo de la moneda las probabilidades de los tipos de secuencia difieren mucho de las probabilidades de las secuencias concretas. Esto es debido a que también un estado macroscópico puede darse en forma de un gran número de estados microscópicos. En un sistema físico, el equivalente a la curva azul de la figura 1 es el número de estados microscópicos compatibles con un estado macroscópico dado. Boltzmann demostró que la entropía, una magnitud que había aparecido al estudiar gases y motores térmicos pero cuya naturaleza era aún misteriosa, estaba relacionada con ese número de estados microscópicos compatibles. En concreto, la entropía es proporcional al logaritmo de dicho número, y la importancia de esta relación matemática es tal que se encuentra grabada sobre la tumba de Boltzmann en el cementerio de Viena.

Pensemos, por ejemplo, en un gas contenido en un recipiente, a su vez dividido en dos por una pared permeable. La energía es menor en el recinto inferior. Por lo tanto, cada molécula "preferirá" estar en ese recinto, igual que con nuestra moneda cada tirada "prefería" ser cruz en lugar de cara. Sin embargo, en una observación macroscópica no nos interesa el comportamiento de cada una de las moléculas sino sólo el comportamiento colectivo, que en este caso podría ser la fracción de moléculas que haya en cada recinto. El problema del gas es entonces idéntico al problema de la moneda, y de la misma manera que lo hicimos con las 200 tiradas, se puede demostrar que en el estado macroscópico más probable habrá una fracción apreciable de moléculas



2. Un gas cuyas moléculas pueden estar en el recinto superior o en el inferior. En el inferior la energía es menor pero, debido al efecto de la entropía, el estado macroscópico más probable es aquel en el que una fracción apreciable de moléculas se encuentran en el recinto superior

en el recinto superior. El valor concreto de esta fracción es el resultado de la combinación de un factor energético y de un factor entrópico, que son los análogos a las curvas verde y azul, respectivamente, de la figura 1. Los detalles matemáticos, aunque no son muy complicados, van más allá de la intención de este artículo. Pero sí es interesante mencionar que, cuanto menos pesadas son las moléculas y más calientes están, más importante es el factor entrópico y, por tanto, las moléculas tienden a estar repartidas por igual en los dos recintos. Fíjense también que, si no fuera por el factor entrópico, el estado más probable sería aquel en el que todas las moléculas se encontrarían en el recinto inferior, con la mínima energía posible. De hecho, así ocurre cuando la temperatura es de menos 273 grados centígrados. Si no fuera por la entropía, las moléculas que componen el aire que respiramos se precipitarían al suelo. La entropía también explica por qué moléculas muy ligeras, como las de hidrógeno, no pueden ser retenidas por el campo gravitatorio terrestre y no forman parte de la composición del aire.

parr-km0@zenon.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

La misteriosa ley del primer dígito

La ley del primer dígito, también conocida como *ley de Benford*, es un fenómeno matemático sorprendente, casi mágico, que revela una cierta universalidad en los aspectos cuantitativos del mundo, una propiedad común a series de datos que provienen de fuentes muy diversas.

Para entenderla es mejor relatar cómo fue descubierta. El descubridor de la "ley" no fue Benford, sino un matemático y astrónomo del siglo XIX llamado Newcomb. En 1880 no existían calculadoras ni ordenadores, pero los científicos e ingenieros realizaban cálculos complejos gracias a los logaritmos. Para multiplicar dos números, por ejemplo, basta averiguar sus logaritmos, sumarlos y encontrar el *antilogaritmo* de esta suma. Por ello, los científicos consultaban constantemente las *tablas de logaritmos*. Las tablas son como un diccionario en el que las entradas fuesen números ordenados, no de menor a mayor, sino de forma parecida a la alfabética: primero los que comienzan por 1, luego los que comienzan por 2, y así sucesivamente.

Newcomb se dio cuenta de que las primeras páginas de las tablas de la biblioteca de su facultad estaban mucho más sucias y manoseadas que las últimas. ¿Quizás en su facultad era más corriente consultar el logaritmo de los números que comienzan por 1 y por 2 que el de aquellos que comienzan por 9? Fue a la biblioteca de una facultad diferente a la suya, a escuelas de ingenieros, de matemáticos, de físicos, de astrónomos,... y el resultado fue el mismo: los números que comienzan por 1 eran más consultados que los que comienzan por 9.

El propio Newcomb encontró, mediante un argumento poco riguroso, una fórmula para la probabilidad de que un número en una serie de datos comience por el dígito d. La fórmula es:

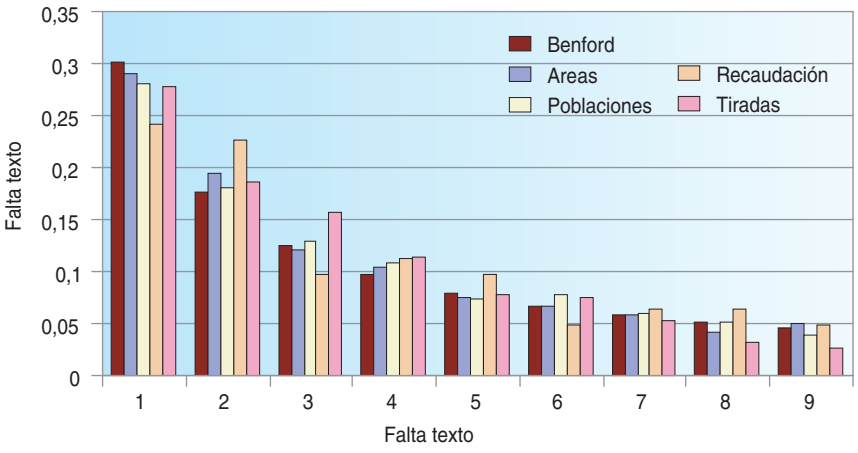
$$p_d = \log_{10}(1 + 1/d)$$

Utilizando esta fórmula se encuentra que, para $d = 1$ la probabilidad es, aproximadamente, del 30 %, para $d = 2$ del 17,6 %, para $d = 3$ del 12,5 %, para $d = 4$ del 9,7 %, para $d = 5$ del 7,9 %, para $d = 6$ del 6,7 %, para $d = 7$ del 5,8 %, para $d = 8$ del 5,1 % y para $d = 9$ del 4,6 %.

La observación de Newcomb no tuvo demasiada repercusión hasta que otro matemático, Frank Benford, publicó en 1932 un artículo con más de 20.000 números extraídos de multitud de fuentes: áreas de ríos, poblaciones de ciudades, números que aparecen en la primera página de los periódicos, constantes físicas, pesos atómicos... Contó con paciencia el primer dígito de todos esos datos, calculó las probabilidades para cada dígito y demostró que la mayoría de estos conjuntos de datos verifican la fórmula de Newcomb. Por eso ahora conocemos la ley con el nombre de Benford.

Sin embargo, obviamente hay conjuntos de números o de datos que no verifican la ley: en los números que han salido en la lotería todos los dígitos deben tener una probabilidad del 10 %; los pesos en kilogramos de los europeos adultos se encuentran alrededor de 80 kilogramos y por tanto el 8, el 7 y el 9 serán mucho más probables que el 2 o el 3.

Entonces, ¿qué datos verifican la ley de Benford? La respuesta no está muy clara. Por eso entrecomillo la palabra "ley". No hay una caracterización precisa de las series de datos que verifican la ley, pero lo cierto, y lo sorprendente, es que la mayoría de los datos que se pueden extraer de fenómenos naturales, sociales y económicos sí la verifican. Yo mismo he comprobado la ley para datos que provienen de fuentes muy diversas: el área y la población de los países del mundo, la recaudación de las taquillas de los cines de cada provincia del estado español durante 1997, y la tirada de las revistas y periódicos de nuestro país también durante 1997. Aunque hay algunas discrepancias, el acuerdo es impresionante, sobre todo teniendo en cuenta el origen tan variado de los datos. De hecho, las cuatro series



La frecuencia del primer dígito de varias series de datos comparadas con la ley de Benford. Las series de datos son: áreas y poblaciones de 242 países del mundo, recaudación de los cines de 62 provincias y comunidades autónomas; tirada de 413 publicaciones

que se presentan en la figura se escogieron completamente al azar de varias bases de datos que tenía a mi disposición a través de Internet o de anuarios de periódicos en formato digital.

¿Cómo es posible que datos tan distintos compartan esta extraña propiedad? Aunque no definitiva, una de las respuestas más convincentes se basa en la llamada *invariancia bajo cambio de escala* o *autosimilaridad*.

Imagine que el tamaño de su cuerpo se reduce a la mitad, como si usted se reencarnara en el protagonista de *El increíble hombre menguante*. La altura de sillas y mesas y el tamaño de cualquier objeto se multiplicaría por dos y, como consecuencia, su vida cotidiana resultaría bastante más difícil. No hay ninguna duda de que el mundo no sería el mismo si todas las longitudes, salvo una que sirva de referencia, se multiplicasen (o se dividiesen) por dos. Sin embargo, en ciertos casos, en ciertos "mundos", no ocurre así. Si algo no cambia cuando las distancias pertinentes se multiplican por dos (o por otra cantidad), decimos que es *autosimilar* o *invariante bajo cambio de escala*. También decimos que es un *fractal*. Aunque no lo parezca, hay muchos fractales en la naturaleza, sobre todo en formaciones geológicas, en fenómenos relacionados con fluidos, en seres vivos y en sistemas sociales y económicos.

¿Qué tienen que ver los fractales con la ley de Benford? Imagine un "mundo fractal", un mundo que no cambia si las distancias se multiplican por dos. Si recogemos datos que corresponden a longitudes, la distribución de estos datos tiene que ser la misma si los multiplicamos por dos. O, para expresarlo de forma más sencilla, si escribimos una tabla con esos datos y otra tabla con esos mismos datos multiplicados por dos, las dos tablas tienen que ser prácticamente indistinguibles. Por supuesto, no pueden ser completamente indistinguibles, ya que si la primera tabla comprende números entre uno y diez mil, la segunda comprenderá números entre dos y veinte mil. Pero lo importante es que la parte central de ambas tablas tiene que ser similar si, como hemos supuesto, los datos provienen de un objeto fractal.

Ahora bien, los números que empiezan por 1 en la primera tabla empezarán por 2 y por 3 en la segunda; los que empiezan por 2, en la segunda tabla empezarán por 4 y 5, y así sucesivamente. Los que empiezan por 5, 6, 7, 8 y 9 en la primera tabla, tendrán que empezar por 1 en la segunda. Como las dos tablas son indistinguibles, llegamos a la conclusión de que las probabilidades pd de que el primer dígito de los datos sea d deben verificar:

$$p_1 = p_2 + p_3$$
$$p_2 = p_4 + p_5$$
$$p_3 = p_6 + p_7$$
$$p_4 = p_8 + p_9$$
$$p_5 + p_6 + p_7 + p_8 + p_9 = p_1$$

El lector puede comprobar que las probabilidades dadas por la ley de Benford cumplen estas cinco ecuaciones y, repitiendo el argumento con otros factores distintos de dos, se puede comprobar que la ley de Benford es la única completamente invariante bajo cambios de escala arbitrarios.

Si este argumento es el verdadero origen de la ley de Benford, entonces la ley nos estaría indicando que la autosimilaridad es ubicua en la naturaleza y en los fenómenos sociales y económicos. Una de las razones de esta posible ubicuidad es que la autosimilaridad aparece cuando no hay una *magnitud característica* en un sistema. Por ejemplo, en las leyes que rigen la formación de paisajes geológicos, no parece que haya ninguna escala de longitud relevante y por ello dichos paisajes suelen ser autosimilares. Del mismo modo, uno podría decir que no hay una cantidad de lectores característica en la prensa de nuestro país y por ello la tirada de las revistas y periódicos es autosimilar, como indica la figura. Sin embargo, todas estas explicaciones están lejos de ser satisfactorias.

Independientemente de cuál sea su origen, algunos investigadores han intentado encontrar aplicaciones a la ley de Benford. Mark Nigrini propuso, en su tesis doctoral de 1992, utilizar la ley para detectar fraudes fiscales o contables, suponiendo que los números de una declaración fraudulenta no siguen la ley de Benford, mientras que los de una declaración honrada sí lo hacen. El método, que Nigrini denomina *análisis digital* (aunque crearía menos confusión llamarlo *análisis de dígitos*), se ha empleado en varias ocasiones y ya ha sido útil para "pescar" a más de un defraudador. El método descubrió, por ejemplo, que una directiva de una cadena de moteles en Estados Unidos estaba firmando cheques para pagar falsas operaciones de corazón. Por supuesto, una vez que la ley de Benford sea conocida por los posibles defraudadores, el análisis digital servirá de muy poco.

Otra aplicación de la ley de Benford ha sido sugerida por uno de los más conocidos expertos en programación informática: Donald Knuth. La idea de Knuth es que un ordenador tiene que manejar datos que probablemente sigan la ley de Benford. Por tanto, podríamos diseñar ordenadores que sean más rápidos calculando o leyendo de discos duros y memorias RAM aquellos números que empiezan por 1.

Alguien podría haber sugerido en el siglo XIX que las páginas de las tablas logarítmicas correspondientes a números que empiezan por 1 fueran más robustas y se accediera más fácilmente a ellas. Curiosamente, la propuesta de Knuth no es más que una versión moderna de esta idea.

Algunos lectores me han hecho notar que la relación entre el problema de la moneda falsa y la teoría de la información (*véase* Juegos matemáticos de octubre de 2002) ha sido tratada por Juan Domínguez en un artículo de la revista *Qüestió* (junio de 1983). Domínguez relaciona el problema con el de la búsqueda de bits erróneos en un mensaje, y hace notar que con tres pesadas se puede detectar la moneda falsa en un conjunto de 12 incluso añadiendo la posibilidad de que ninguna de las monedas sea falsa. Con ello el número total de posibilidades es $12 \times 2 + 1 = 25$, aún menor que el número de posibles resultados, que es $3 \times 3 \times 3 = 27$. Finalmente, el número de pesadas necesarias para resolver el problema con n monedas debe ser mayor que el logaritmo en base tres de $2n$.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

El número de oro

El número de oro, $\phi = (1+\sqrt{5})/2 = 1{,}6180339887499...$, es uno de los protagonistas indiscutibles de ese misterioso espacio en donde se encuentran matemáticas, arte y ciencia. Es un número singular desde el punto de vista matemático pero también estético, al menos según el canon iniciado por los griegos. Los *rectángulos de oro* o *áureos*, es decir, rectángulos en los que el cociente entre el lado mayor y menor es el número de oro, se hallan por doquier en el Partenón, en templos y construcciones griegas y en la composición de edificios, cuadros y fotografías a lo largo de toda la historia del arte.

Los rectángulos de oro están caracterizados por una interesante propiedad de autosimilaridad. Si de un rectángulo áureo extraemos un cuadrado cuyo lado coincida con el lado menor, el rectángulo que queda es también áureo, como ocurre con el rectángulo verde de la figura 1. Podríamos extraer un nuevo cuadrado y quedarnos con un rectángulo áureo más pequeño, y así sucesivamente, encontrando siempre rectángulos áureos.

Numéricamente, esta autosimilaridad equivale a lo siguiente:

$$\frac{1}{\phi} = \frac{\phi - 1}{1}$$

La ecuación anterior puede también escribirse en la forma:

$$\phi = 1 + \frac{1}{\phi}$$

que nos indica que el desarrollo decimal del inverso del número de oro es igual al desarrollo del propio número, es decir $\phi = 1{,}618033...$ y $1/\phi = 1{,}618033...$

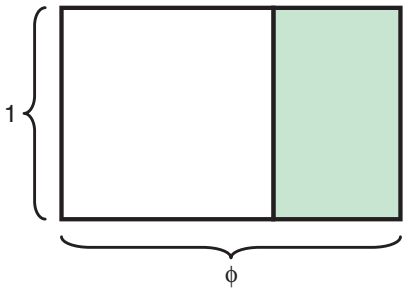
Pero, al margen de estas curiosidades matemáticas, ¿es este rectángulo el "más placentero a la vista", el más armonioso, como se supone que afirma el canon clásico? Probablemente no se pueda saber si realmente es o no más placentero, y, en caso de serlo, cuánto de esa sensación es producto de nuestra cultura. Sin embargo, en mi opinión, hay una cualidad de los rectángulos áureos que se puede apreciar en la fachada de los templos griegos. Se trata de una especie de referencia a alguna medida humana que hace que el edificio, aun siendo monumental, se muestre asequible, cercano. Quizás esta sensación provenga no tanto de la forma del rectángulo, como de que su proporción se repita en todas las escalas, desde los capiteles y los altares hasta la planta del templo completo. En cualquier caso, creo que cualquiera que se haya aproximado a uno de estos templos habrá podido percibir esta combinación de grandeza y accesibilidad.

Todas estas cuestiones han sido el objeto, e incluso la obsesión, de muchos y voluminosos estudios que entrecruzan matemáticas e historia del arte. En Internet pueden también encontrar cientos de páginas dedicadas a este tema. Nosotros vamos a ser más modestos y nos centraremos en algunas de las propiedades matemáticas y geométricas de ϕ. De la última de las ecuaciones anteriores, se deduce que el número de oro puede escribirse de esta curiosa manera:

$$\phi = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ldots}}}}$$

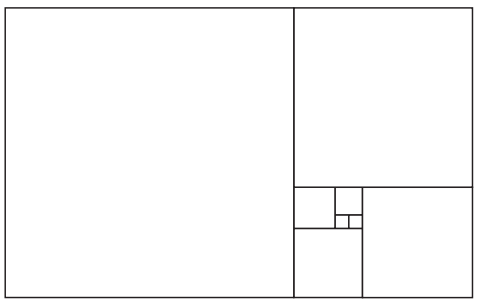
Esta forma de escribir un número como quebrados encabalgados unos sobre otros se llama *desarrollo en fracciones continuas*. En general, el desarrollo de cualquier número x es:

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots}}}$$

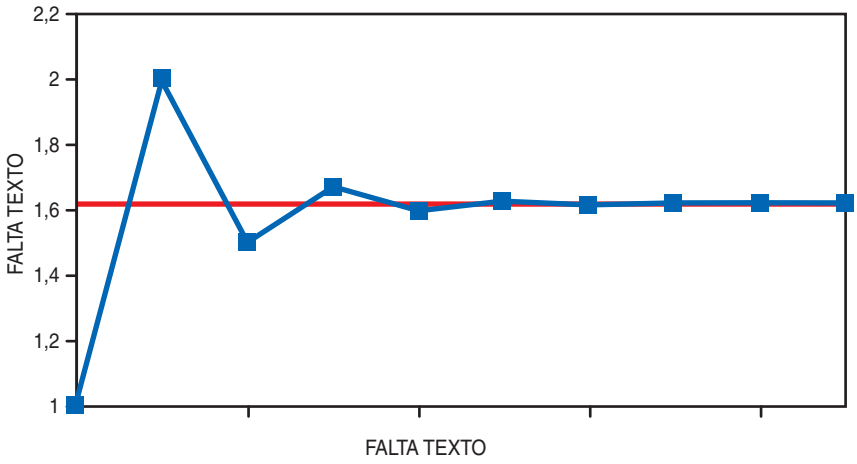


1. Si de un rectángulo áureo, de lados 1 y ϕ, retiramos el cuadrado blanco de lado 1, el rectángulo resultante (en verde) *es también áureo*

y se suele escribir $x = [a_0; a_1, a_2, a_3,...]$. El desarrollo en fracciones continuas es una forma de representar números tan válida como nuestro sistema decimal o como la representación binaria que utilizan los ordenadores. Todas ellas son formas de escribir números. En algunas de ellas ciertas operaciones son más sencillas que en otras. Por ejemplo, la suma es fácil de realizar en el sistema decimal o en el binario. En el decimal es muy sencillo multiplicar o dividir por 10. Sin embargo, en los desarrollos en fracciones continuas es muy simple, por ejemplo, hallar el inverso de un número. En efecto, el lector puede comprobar que, si $x = [a_0; a_1, a_2,$



2. Construcción de un rectángulo áureo a partir de cuadrados cuyos lados siguen la serie de números de Fibonacci: dos cuadrados de lado 1, otro de lado 2, otro de lado 3, 5, 8, 13 y 21



3. Los puntos azules son el resultado de dividir un número de Fibonacci por el inmediato anterior: 1/1 = 1, 2/1 = 2, 3/2 = 1,5, etc. La línea roja representa el valor de ϕ. El octavo Fibonacci divido por el séptimo, 21/13 = 1,615 es ya indistinguible de ϕ en la gráfica

a_3,...], siendo a_0 distinto de cero, entonces $1/x = [0; a_0, a_1, a_2, a_3,...]$.

Cada representación hace más evidentes ciertas particularidades de los números. En el caso del número de oro, su singularidad se pone de manifiesto en su desarrollo en funciones continuas, que es $\phi = [1;1,1,1,...]$. Esta singularidad se debe a que los desarrollos en fracciones continuas están relacionados con la forma de dividir un rectángulo en varios cuadrados. Un método gráfico para encontrar la representación en fracciones continuas de un número x, mayor que uno, consiste en construir un rectángulo de base x y de altura 1 y extraer de él tantos cuadrados de lado 1 como sea posible. El número de cuadrados será a_0. El rectángulo que queda (de lados 1 y $x - a_0$) se gira para que el lado mayor sea el horizontal y se vuelve a repetir el procedimiento de extraer cuadrados, en este caso de lado $x - a_0$. El número de cuadrados será a_1. El procedimiento se repite para encontrar a_2, a_3, etc. Si este procedimiento se comienza con un rectángulo áureo, el número de cuadrados que se pueden extraer es uno y el rectángulo que queda es de nuevo áureo, lo cual explica que el desarrollo en fracciones continuas de ϕ sea una sucesión infinita de unos.

Otra forma interesante de llegar al número de oro es utilizando los llamados *números de Fibonacci*, F_k. Los dos primeros son iguales a uno, $F_1 = F_2 = 1$, y el resto se construye como la suma de los dos anteriores, $F_k = F_{k-1} + F_{k-2}$. Es decir, $F_3 = 1 + 1 = 2$, $F_4 = 2 + 1 = 3$, $F_5 = 3 + 2 = 5$, $F_6 = 8$, $F_7 = 13$, $F_8 = 21$, $F_9 = 34$. Resulta que el cociente entre dos números de Fibonacci consecutivos se acerca cada vez más al número de oro, como puede verse en la figura 3.

Existe una razón geométrica para este acercamiento, que se ilustra en la figura 2, y está también basada en rectángulos que contienen cuadrados. Construimos un cuadrado de lado unidad y otro igual pegado a su lado izquierdo. Sobre ellos construimos un nuevo cuadrado cuyo lado sea 2, la suma de los dos lados anteriores. A la izquierda de este conjunto formado por tres cuadrados, construimos uno nuevo, en este caso de lado 3. Vamos anexionando a la figura nuevos cuadrados cuyos lados siguen la secuencia de Fibonacci. Pues bien, los rectángulos resultantes son cada vez más parecidos a un rectángulo áureo. Al lector no le será muy difícil adivinar por qué, comparando la propiedad de autosimilaridad del rectángulo áureo con este procedimiento de construcción. Podrá también comprobar que el desarrollo en fracciones continuas de F_{k+1}/F_k consiste en $k - 2$ unos seguidos de un 2, seguido de infinitos ceros.

Al igual que el número de oro, los de Fibonacci también aparecen en muchos diseños de la naturaleza. Por ejemplo, el número de pétalos de la mayoría de las flores coincide con uno de los de la serie. Se han utilizado también en música, en pintura y en arquitectura. Virgilio, Leonardo da Vinci, Beethoven, Bach, Debussy, Bartok, Le Corbusier, el arquitecto del Partenón, todos ellos han utilizado ϕ o la serie de Fibonacci. En poesía la proporción áurea se ha utilizado para estructurar largos poemas y algunas métricas están basadas en el 8 y el 13, que son números de Fibonacci. En el cuadro les presento un pequeño ejercicio, casi un experimento: un poema cuyos versos tienen un número de letras dado por la serie de Fibonacci. El poema hace además referencia al propósito original de Fibonacci cuando introdujo sus números: describir cómo crece la población de una cierta especie en la que cada individuo se reproduce dos veces a lo largo de su vida.

Debido al crecimiento de la serie de Fibonacci, los versos dibujan una suave pendiente exponencial y las proporciones entre las longitudes de versos consecutivos se acercan cada vez más al número de oro. ¿Se consigue con ello crear una cierta armonía en el poema?, ¿una evocación de los modos en que la naturaleza realiza sus propios diseños?

y
a
su
vez
salen
del árbol
pisan las hojas
secas del camino y cantan
extraños sones murmullos de luz y de agua
se colman de semillas y se vuelven dorados y se abandonan al viento
...

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Números y palabras

Para la mayoría de la gente los números y las palabras pertenecen a reinos separados e incluso antagónicos. O se es "de letras", o "de ciencias", y cada materia pertenece a uno u otro campo sin posibilidad de conexión entre sí. Sin embargo, desde hace varias décadas, existe una disciplina que aplica técnicas matemáticas al estudio del lenguaje: la *lingüística matemática*. Una parte importante de esta disciplina es la *lingüística cuantitativa*, que consiste en el estudio estadístico de textos, facilitado por la mayor potencia y capacidad de memoria de los ordenadores actuales.

Vamos a explorar algunas curiosidades de esta lingüística cuantitativa utilizando textos que se pueden encontrar en Internet: *Don Quijote*, *Cien años de soledad* y el original en inglés del *Ulysses* de Joyce. Los he analizado por medio de *TextStat*, un programa gratuito que realiza estadísticas elementales de cualquier texto, creado por la Universidad Libre de Berlín. Se pueden encontrar muchos programas como éste en Internet, algunos gratuitos y otros comerciales, bajo el nombre genérico de *Natural Language Processing (NPL) software* (programas para el procesamiento del lenguaje natural). Han sido diseñados para tratar aspectos gramaticales del lenguaje, para administrar grandes conjuntos de textos que se denominan *corpus*, para encontrar *concordancias*, es decir, las apariciones de una determinada palabra en un corpus, y para algunas otras funciones relacionadas con el estudio matemático del lenguaje.

Uno de los primeros hallazgos de la lingüística cuantitativa fue la llamada *ley de Zipf*, una sorprendente regularidad en cómo se distribuyen las palabras en un texto de cualquier lengua. Se toma un texto suficientemente largo y se cuenta el número de veces que aparece en él cada palabra. Hay palabras comunes, como "el", "de" o "que", que aparecerán un gran número de veces, y otras más raras que sólo aparecerán una vez. A continuación colocamos las palabras en una lista, ordenándolas de más a menos frecuentes. El orden que una palabra ocupa en la lista se denomina *rango*. Así, en el caso de *Cien años de soledad*, el rango de "de" es 1, el de "la" es 2, el de "que", 3, el de "y", 4, el de "el", 5, etc. Pues bien, la ley de Zipf afirma que la frecuencia f de una palabra dada es inversamente proporcional a su rango r, es decir:

$$f = a/r$$

en donde a es una constante que depende del texto utilizado. Se trata evidentemente de una ley aproximada, puesto que puede dar frecuencias no enteras e incluso menores de uno para rangos muy grandes. Una generalización de la ley, también aproximada pero que se adapta mejor a cualquier tipo de texto, supone que la distribución de frecuencias es una *ley de potencias*:

$$f = a/r^b$$

en donde b es un exponente cercano a 1.

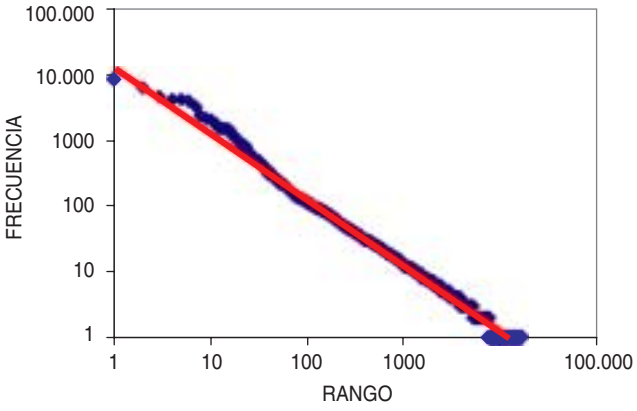
Para ver si un texto satisface la ley de Zipf o su generalización, es necesario construir la lista y representar las frecuencias f de cada palabra en función de su rango r. Sin embargo, es más útil hacer esa representación en *ejes logarítmicos*, es decir, en ejes en donde cada cantidad no varía de unidad en unidad sino en intervalos dados por las distintas potencias de diez. Esta representación equivale a dibujar la gráfica del logaritmo de f frente al logaritmo de r, que se relacionan de la siguiente forma:

$$\log f = \log a - b \log r$$

Como a es una constante, esta ecuación nos dice que la gráfica de $\log f$ en función de $\log r$, o, equivalentemente, la de f en función de r con ejes logarítmicos, será una recta inclinada hacia abajo y con pendiente b.

Veamos si *Cien años de soledad* verifica la ley de Zipf. La novela tiene un total de 138.014 palabras, entre las cuales hay 16.019 diferentes. TextStat realiza una tabla con la frecuencia de cada una de estas 16.019 palabras.

La lista contiene algunas curiosidades y enigmas. Por ejemplo, el artículo femenino "la" aparece significativamente más veces que el masculino "el", mientras que el plural "los" aparece más veces que "las" y el indeterminado "un" aparece más veces que "una". Hay una primera explicación bastante sencilla: "la", a diferencia de "el", no es sólo artículo determinado sino también pronombre. Aun así, la explicación no es completa porque, más abajo en la lista, comprobamos



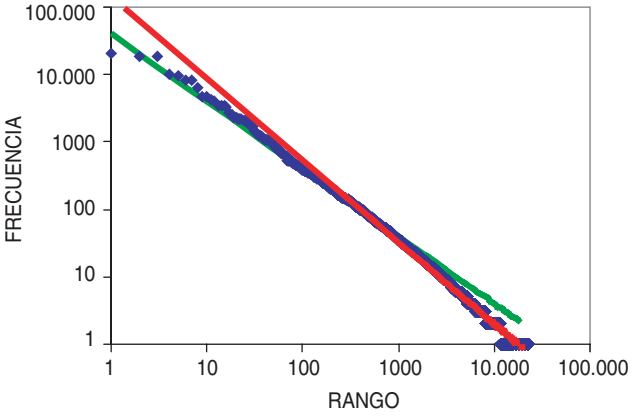
1. La ley de Zipf para Cien años de soledad

que “lo” aparece 888 veces que, sumadas a las 4054 veces de “el”, aún se encuentran muy por debajo de las apariciones de “la” (6110). Sigue siendo curioso que los dos artículos indeterminados tengan casi la misma frecuencia mientras que el plural “unos” sale 42 veces y “unas” sólo 20. Lo mismo ocurre con otros pronombres y adjetivos. Por ejemplo, “ese” sale 75 veces y “esa” 105, mientras que “esos” sale 30 y “esas” sólo 11. “Otro” sale 106, “otra” 93 veces, “otros” 55 y “otras” 21. En los plurales las diferencias siempre son bastante significativas. Se podría pensar que es debido a que en castellano se toma el masculino para plurales en donde hay tanto mujeres como varones, pero, analizando las distintas apariciones de “otros” y “otras”, por ejemplo, se puede comprobar que este argumento no da cuenta de la diferencia. En mi opinión, una posible explicación se encuentra en el hecho de que existe un gran número de sustantivos femeninos en castellano que se refieren a cualidades, como “blancura” o “pereza”, y cuyo plural es inexistente o muy raro.

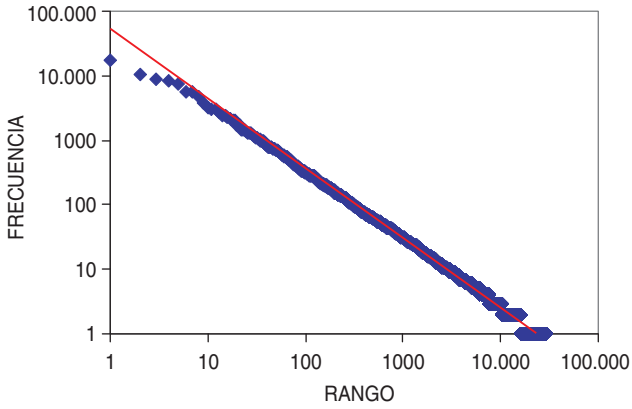
Volviendo a la ley de Zipf, en la figura 1 pueden ver la gráfica de la frecuencia en función del rango. Los puntos azules son las frecuencias de cada palabra en el texto y la línea roja es la recta que mejor se aproxima a los puntos: una ley de potencias con $a = 13.083$ y $b = 1{,}0086$. El exponente b es muy cercano a 1, de modo que podemos decir que la ley de Zipf original se cumple con bastante aproximación en la novela de García Márquez.

En la figura 2 podemos ver la ley de Zipf para *Don Quijote*, en el que hay 22.941 tipos de palabras entre las 381.222 que componen el texto. La línea roja representa de nuevo la ley de potencias que mejor se ajusta a los datos “experimentales” y en este caso es la función $f = 142.271\ r^{-1{,}2136}$, es decir, una ley de potencias con exponente $b = 1{,}2136$. El exponente se aleja significativamente de 1, el exponente de la ley de Zipf original, pero también podemos observar que no reproduce bien el comportamiento de las palabras más frecuentes. He dibujado en verde la ley de Zipf “genuina”, es decir, con exponente $b = 1$, que mejor se ajusta a los datos y el resultado no es peor que el de la mejor ley de potencias. Lo que ocurre en este caso es que la ley de Zipf genuina se ajusta bien a los puntos con mayor frecuencia, pero se desvía de los puntos con mayor rango y menor frecuencia. ¿Qué se puede concluir entonces? Yo diría que ni la ley de Zipf ni la de potencias se ajustan a los datos de *El Quijote*. Aunque las palabras más frecuentes sí cumplen aquélla, parece haber un número demasiado pequeño de palabras “raras”, es decir, de palabras que aparecen sólo una vez. Esta escasez hace que al ajustar una ley de potencias, ésta se incline en exceso y no pueda dar cuenta de modo preciso del comportamiento de las palabras más frecuentes.

Para ver la universalidad de la ley de Zipf, les presento el análisis del original en inglés del *Ulises* de James Joyce. A pesar de ser una novela en donde hay una mayor experimentación con el lenguaje, los datos se ajustan bastante bien a la ley de Zipf. Frecuencia y rango están relacionados por una ley de potencias



2. Ley de Zipf para el Quijote. *La línea roja es la ley de potencia que mejor se ajusta a los datos (puntos azules), mientras que la línea verde es una ley de Zipf “genuina” con exponente* b = *1*



3. La ley de Zipf para el original en inglés del Ulises *de Joyce*

con $a = 52.467$ y exponente $b = 1{,}0793$, aunque se observa una desviación con respecto a la ley en las palabras más frecuentes.

Zipf introdujo su ley en 1949, en un libro titulado *El comportamiento humano y la ley del mínimo esfuerzo*. La razón de este título es que la ley puede derivarse suponiendo que el lenguaje natural se ha desarrollado de modo que transmita la mayor cantidad de información con el menor número de palabras. Benoit Mandelbrot también realizó en 1951 una demostración similar. Sin embargo, ambas demostraciones implican una relación entre el rango de la palabra y su longitud, de modo que las palabras más frecuentes son las más cortas. Esta relación es cierta, como se ve en el caso de *Cien años de Soledad*, cuyas palabras más frecuentes son de una o dos sílabas, pero deja de ser cierta para palabras de mayor rango.

Otra herramienta que sin duda atraerá al curioso por el lenguaje es la consulta del corpus diacrónico de la Real Academia Española, que puede realizarse por Internet (www.rae.es, Consulta Banco de datos) y en el que se pueden buscar palabras y expresiones y observar su utilización desde el origen de nuestra lengua hasta nuestros días.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

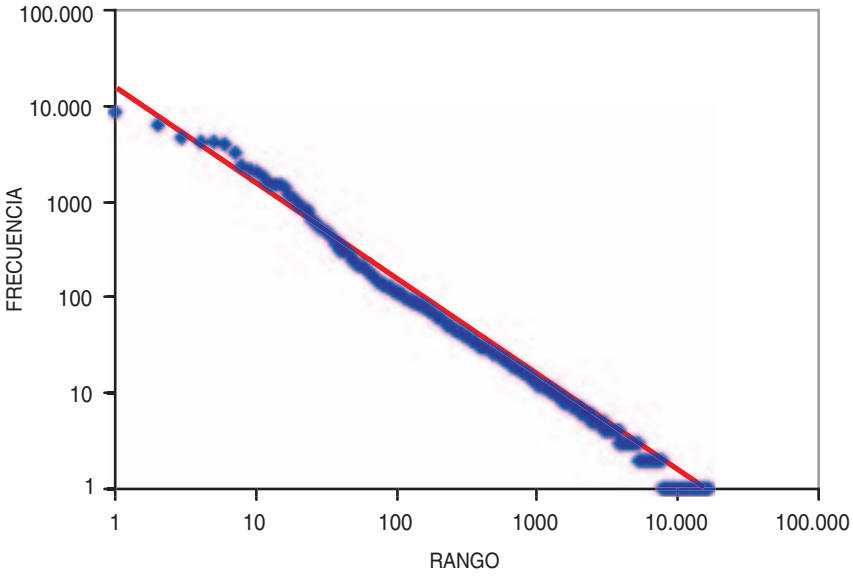
Más sobre números y palabras

En el último artículo de *Juegos Matemáticos* exploramos una sorprendente regularidad en la estadística de un texto literario: la ley de Zipf. La ley nos dice que, si ordenamos las palabras que aparecen en un texto de más a menos frecuentes, la frecuencia con la que aparece una palabra en el texto es inversamente proporcional al puesto que ocupa en la lista. Aunque no lo comentamos entonces, la constante de proporcionalidad es aproximadamente igual al número de palabras diferentes que aparecen en el texto. Es decir, la frecuencia de una palabra es:

$$f(r) = \frac{V}{r}$$

en donde V es el vocabulario del texto, es decir, el número de palabras *diferentes* que aparecen en él, y r el rango de la palabra, o lugar que ocupa en la lista en la que todas las palabras del texto se ordenan de más a menos frecuentes.

Veamos un ejemplo. En la novela de Gabriel García Márquez *Cien años de soledad*, que tiene un vocabulario de 16019 palabras, la palabra "de" es la más común y aparece 8684 veces. Como es la palabra más común, su rango es 1 y la fórmula de la ley de Zipf predice para "de" una frecuencia igual a 16019. Esta cifra es casi el doble de la frecuencia real de la palabra por lo que parece, en principio, que la ley falla estrepitosamente. Si la aplicamos a las siguiente palabras de la tabla, vemos que la discrepancia es menor pero sigue siendo considerable (*véase la tabla 1*). Sin embargo, esas discrepancias no son tan grandes como aparentan si uno observa la *figura 1*, en donde se dibujan las frecuencias de cada palabra (*puntos azules*) junto con la predicción de la ley de Zipf (*línea roja*).



1. Los puntos azules muestran la relación entre la frecuencia y el rango de todas las palabras que aparecen en Cien años de soledad. La línea roja es la ley de Zipf. Tanto la escala vertical como la horizontal son logarítmicas, lo cual enmascara ligeramente las discrepancias entre la ley y los datos reales

Existen algunas modificaciones de la ley de Zipf que se aproximan mejor a las frecuencias reales de las palabras. Comentamos algunas de ellas en al artículo del mes pasado. Sin embargo, la ley de Zipf original, aunque es sólo válida de modo aproximado, permite hacer algunas predicciones interesantes acerca de la estadística de un texto. Una de ellas es la relación entre el tamaño de un texto o número total de palabras que lo componen, y su vocabulario. Esta relación nos da una idea de la riqueza de vocabulario de cada texto. El tamaño N de un texto se puede obtener sumando todas las frecuencias del vocabulario:

$$N = f(1) + f(2) + f(3) + \ldots + f(V)$$

Si las frecuencias verifican la ley de Zipf, entonces la suma es:

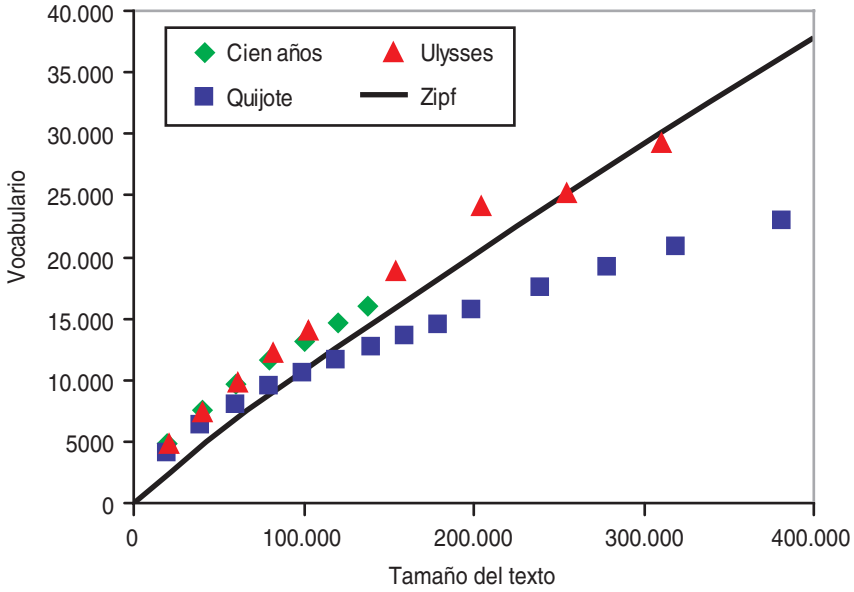
$$N = V\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{V}\right)$$

La suma entre paréntesis se puede realizar de modo aproximado, cuando el vocabulario es muy grande (al lector con algunos conocimientos de matemáticas superiores no le resultará difícil entender que esta aproximación consiste en sustituir la suma por una integral de la función $1/r$). El resultado es

$$N = V \ln V$$

en donde ln es el logaritmo neperiano. Esta sencilla fórmula nos indica la relación entre el tamaño y el vocabulario de un texto. En principio es válida para cualquier texto, siempre que éste cumpla la ley de Zipf. Como esta ley se aplica a cualquier tipo de texto en cualquier idioma, el resultado sorprendente es que todos los textos en todos los idiomas tienen aproximadamente la misma riqueza de vocabulario.

Veamos hasta qué punto es esto cierto. En la figura 2 pueden ver la relación entre el tamaño del texto y el vocabulario para *Cien años de soledad*, *Don Quijote* y la versión inglesa del *Ulysses* de James Joyce. En cada libro vemos cómo el vocabulario va creciendo en función del tamaño del texto. Finalmente, la curva negra es la fórmula que acabamos de deducir a partir de la ley de Zipf. Observen que la fórmula nos da el tamaño a partir del vocabulario, mientras que en la gráfica dibujamos la relación inversa, es decir, el vocabulario en función del tamaño del texto. Sería deseable despejar V en función de N en la ecuación anterior, pero esto no es posible y por lo tanto no disponemos de una fórmula sencilla y compacta que nos diga el vocabulario de un texto de tamaño dado. Sin embargo, esto no es ningún problema para el programa informático que he utilizado para dibujar la curva negra de la figura 2 a partir de la fórmula $N = V \ln V$.



2. Cómo crece el vocabulario en tres conocidos textos en función del tamaño o número total de palabras. La curva negra representa el resultado derivado a partir de la ley de Zipf

De nuevo, a pesar de que existen ciertas discrepancias, la fórmula nos ofrece una descripción aproximada de cómo aumenta el vocabulario en cada texto. La novela que más se aleja del resultado teórico es El Quijote, cuyo vocabulario es menor del que se espera, es decir, en la segunda parte de la novela aparecen menos palabras nuevas de las que predice la ley de Zipf.

Observen que el crecimiento del vocabulario no es proporcional al tamaño del texto. En principio uno podría pensar que en un texto la fracción de palabras distintas, es decir, el vocabulario dividido por el tamaño, es constante. Sin embargo, no puede ser así porque, cuanto mayor es el texto, más difícil es encontrar palabras que no hayan aparecido con anterioridad. De hecho, la fórmula del logaritmo neperiano tiene una propiedad que no pueden verificar los textos reales. El vocabulario crece indefinidamente con el tamaño del texto, lo cual es imposible porque el número total de palabras de una lengua, incluyendo todas las variantes morfológicas —plurales, conjugaciones verbales, etc.— es finito. Quizás está finitud en el léxico de la lengua es lo que hace que el vocabulario de El Quijote no alcance las predicciones de la ley de Zipf. Algunos investigadores han desarrollado fórmulas más complejas que tienen en cuenta estas limitaciones y se adaptan a todo tipo de texto. Con ellas deducen también relaciones entre el vocabulario y el tamaño, utilizando técnicas matemáticas sofisticadas. Como ven, la estadística es capaz de encontrar y cuantificar regularidades en algo tan humano como es el uso del lenguaje, e incluso en su vertiente más personal y creativa: la literatura.

RANGO	PALABRA	FRECUENCIA	LEY DE ZIPF
1	**de**	8684	16.019
2	**la**	6110	8010
3	**que**	4679	5340
4	**y**	4147	4005
5	**el**	4054	3204
6	**en**	3880	2670
7	**a**	3162	2288
8	**los**	2373	2002
9	**se**	2142	1780
10	**con**	1983	1602
11	**un**	1785	1456
12	**las**	1535	1335
13	**una**	1505	1232
14	**por**	1465	16.019

Palabras más frecuentes en Cien años de soledad

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

Fluctuaciones fatales

Disponemos de un cultivo de bacterias que crece por división de cada uno de los individuos que lo forman. Supongamos que, de un día para otro, cada bacteria tiene una probabilidad 1/5 de morir, una probabilidad 3/5 de duplicarse y una probabilidad 1/5 de seguir viva sin reproducirse. Si inicialmente hay 1000 bacterias en la colonia, aproximadamente 200 morirán, 600 se duplicarán y 200 se mantendrán vivas sin reproducirse. Al día siguiente tendremos, por tanto, 1400 bacterias vivas. Es decir, en media, la población crece un 40 % cada día: el segundo día habrá 1960 individuos, el tercer día 2744, y así sucesivamente. La población crece exponencialmente si nuestras suposiciones acerca de la reproducción y muerte se siguen manteniendo. En diez días, por ejemplo, la colonia alcanzaría casi los 30 mil individuos y en doscientos días llegaría hasta 10^{32}. Este número es demasiado grande para un cultivo real, lo cual indica que las suposiciones no se pueden mantener durante todo el tiempo debido a que la competencia entre las bacterias por los nutrientes o simplemente por el espacio dentro del cultivo hará sin duda disminuir las probabilidades de reproducción y aumentar las de muerte. En consecuencia, nuestro modelo simplificado sólo puede ser válido para poblaciones pequeñas y para los primeros días de evolución de la colonia. Aun así, su análisis revelará algunos efectos interesantes.

Como acabamos de ver, la población de la colonia de bacterias crece, en media, día tras día. La indicación "en media" es aquí muy relevante. "En media" significa que sólo aproximadamente 1/5 de las bacterias morirán de un día para otro y sólo aproximadamente 3/5 se reproducirán, igual que sólo aproximadamente 1/6 de las veces obtenemos un cinco cuando lanzamos un dado no trucado. Cuantas más veces lanzamos el dado, más se acercará la fracción de cincos a 1/6. No obstante, siempre habrá algunas desviaciones con respecto a 1/6 (que no es otra cosa que la probabilidad de que salga un cinco). Estas desviaciones suelen llamarse "fluctuaciones".

Las fluctuaciones son menores cuanto mayor es el número de tiradas del dado y, en el caso de la colonia, serán menores cuanto mayor sea la población de bacterias. Así, si inicialmente el cultivo dispone de 1000 bacterias, las fluctuaciones no serán muy importantes: morirán unas 200 y se reproducirán unas 600. Las pequeñas desviaciones con respecto a estos valores no cambiarán demasiado el comportamiento del cultivo, que crecerá, como hemos visto, aproximadamente en un 40 % cada día. En pocas palabras: en poblaciones grandes, los valores medios son casi iguales a los valores reales. Por el contrario, en poblaciones pequeñas las fluctuaciones pueden ser de vital importancia.

¿Qué ocurre si, inicialmente, disponemos de una única bacteria? Al día siguiente nos encontraremos que nuestro valioso y solitario individuo ha muerto con una probabilidad 1/5 y se ha reproducido con una probabilidad 3/5. Si ha muerto, la colonia se habrá extinguido irremediablemente. Si se ha reproducido, podemos mantener la esperanza de que la colonia crezca en los próximos días, aunque bien pueden estos dos individuos morir al día siguiente, algo que ocurrirá con probabilidad $1/5 \times 1/5 = 1/25$. Es evidente que los primeros días de esta colonia mínima son bastante críticos. Los escasos individuos que la forman son los patriarcas de una familia que, sólo al alcanzar un cierto tamaño, estará libre de desaparecer debido a "fluctuaciones fatales". Esto es lo que probablemente ocurre en la evolución de las especies. Cuando se produce una mutación, sólo un individuo es portador de la misma. Por tanto, aunque la mutación sea ventajosa, su éxito depende de la suerte que corra este único portador de la mutación y sus inmediatos descendientes. Podemos comprobarlo analizando el comportamiento de nuestra colonia. En la figura se dibuja su tamaño en función del tiempo para diferentes simulaciones hechas por ordenador. De las nueve simulaciones, sólo cinco prosperaron, mientras que cuatro se extinguieron en los primeros seis días. Observen que, en estos primeros días, cualquiera de las nueve corre el riesgo de extinguirse.

¿Cuál es la probabilidad de que, partiendo de un solo individuo, la colonia se extinga? En otras palabras, si repetimos el experimento de la figura un gran número de veces, ¿qué fracción de trayectorias acaba cayendo en el eje horizontal? El cálculo de esta *probabilidad de extinción* es un problema muy complicado a primera vista. Si en el primer día la colonia consta de un solo individuo, en el segundo se habrá extinguido con probabilidad 1/5, seguirá con un solo individuo con probabilidad 1/5 o tendrá dos con una probabilidad 3/5. En el tercer día las posibilidades aumentan: la colonia puede ahora constar de cero hasta cuatro individuos. Para calcular la probabilidad de cada una de estas posibilidades hay que tener en cuenta todos los modos en que cada población puede alcanzarse. Por ejemplo, la población de cuatro individuos en el segundo día sólo puede alcanzarse si el patriarca se reprodujo en el primero y segundo día y su vástago lo hizo en el segundo. Eso ocurre con probabilidad $3/5 \times \times 3/5 \times 3/5$. La probabilidad de que la colonia se extinga en el segundo día puede calcularse del mismo modo: o bien el patriarca muere el primer día (probabilidad 1/5), o bien sobrevive al primer día pero muere en el segundo (probabilidad $1/5 \times$

× 1/5) o bien se reproduce el primer día pero tanto él como su hijo mueren en el segundo día (probabilidad 3/5 × 1/5 × 1/5). Por lo tanto, la probabilidad de que la colonia se extinga en el segundo día es

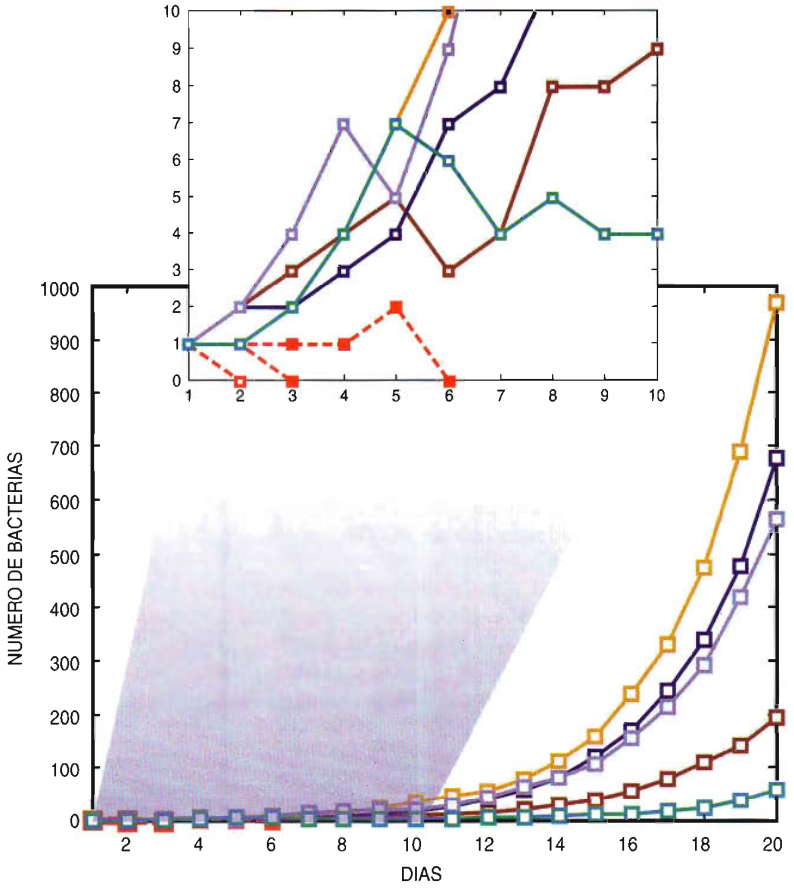
$$q = \frac{1}{5} + \frac{1}{5} \times \left(\frac{1}{5}\right) + \frac{3}{5} \times \left(\frac{1}{5} \times \frac{1}{5}\right) = \frac{33}{125}$$

en donde he colocado entre paréntesis las probabilidades de los eventos que ocurren en el segundo día, por razones que serán evidentes dentro de un momento.

Este modo de calcular la probabilidad de extinción se complica considerablemente al añadir más días, ya que el árbol de posibilidades se va ramificando un día tras otro. Existe, sin embargo, un método bastante ingenioso para calcular la probabilidad de extinción. Llamemos q a esta probabilidad, es decir, a la probabilidad de que se extinga una colonia que inicialmente consta de un solo individuo. ¿Cuál es la probabilidad de que se extinga una colonia que parte de *dos* individuos? Como la "estirpe" de cada individuo no interacciona con la del otro (no hay ni reproducción sexual ni competencia), la probabilidad de que se extingan las dos estirpes será $q \times q$ o q^2 (igual que la probabilidad de sacar un seis doble al tirar dos dados es 1/6 × 1/6).

Volvamos ahora a la colonia que parte de un solo individuo. En su segundo día pueden haber ocurrido tres cosas: A) que el patriarca haya muerto (probabilidad 1/5), B) que siga vivo sin reproducirse (probabilidad 1/5) y C) que se haya reproducido (probabilidad 3/5). Si nos situamos en este segundo día, ¿cuál es la probabilidad de extinción para cada una de estas tres posibilidades? Para A es claramente 1 puesto que la colonia ya está extinta en el segundo día. Para B, la probabilidad de extinción es de nuevo q. Y para C, como hemos visto en el párrafo anterior, la probabilidad de extinción es q^2. La probabilidad q de que se extinga la colonia inicial se puede escribir entonces como una suma que tenga en cuenta estas tres posibilidades:

$$q = \frac{1}{5} + \frac{1}{5} q + \frac{3}{5} q^2$$



Simulación por ordenador de la colonia de bacterias que se describe en el texto. De las nueve veces que se realizó la simulación, en cuatro ocasiones la colonia se extinguió y sólo en cinco continuó creciendo exponencialmente. En la ampliación de los primeros días puede verse, en rojo y línea discontinua, la trayectoria de las colonias extinguidas. Sólo pueden verse tres porque dos de las colonias se extinguieron en el segundo día y sus breves trayectorias coinciden

Ahora tenemos una ecuación bastante simple para q. Lo que en realidad hemos hecho para llegar a esta ecuación es tomar las tres primeras ramas del árbol de posibilidades del que hablábamos antes y darnos cuenta de que cada una se puede tratar como un nuevo árbol de posibilidades. Fíjense que esta ecuación es la misma que la anterior, pero sustituyendo los dos paréntesis por q y q^2, respectivamente.

La solución de la ecuación anterior es $q = 1/3$ ($q = 1$ es también solución, pero no tiene ningún significado ni interés en el problema). Es decir, si hiciéramos un gran número de simulaciones con las probabilidades de muerte y reproducción de nuestra colonia, un tercio de ellas acabarían extinguiéndose. Observen que en el segundo día sólo se extingue 1/5. El resto, 2/15, lo hace en días sucesivos.

Piensen ahora que la colonia de bacterias que consta de un solo individuo puede ser una nueva mutación naciente en la biosfera, el virus solitario que inhalo antes de contraer la gripe e incluso un neutrón que surge al azar en un material radioactivo y puede comenzar una reacción en cadena. En todos estos casos las fluctuaciones iniciales son críticas. Pueden destruir la mutación, la posibilidad de proliferación del virus o hacer que el efecto del neutrón no vaya más allá de unos cuantos destellos de radiación. Son sólo pequeños golpes de azar, pero capaces de cambiar por completo el destino de toda la biosfera.

JUEGOS MATEMÁTICOS

Juan M. R. Parrondo

El examen inesperado y la teoría de juegos

Hace cuarenta años, en marzo de 1963, Martin Gardner describió en esta misma sección de *Scientific American* una célebre paradoja. Un profesor anuncia a sus alumnos que pondrá un examen a las doce en punto de un día de la semana siguiente de manera completamente inesperada, es decir, los alumnos no sabrán el día en que va a realizarse hasta que el profesor lo anuncie a las doce del día elegido. Los estudiantes se van al bar de la facultad y, tras media hora de discusión, brindan convencidos de que es imposible que el profesor logre su propósito. Se han dado cuenta de que el viernes no puede ser el día elegido porque, en ese caso, el jueves por la noche sabrían con certeza que el examen iba a ser el viernes y, por tanto, no sería inesperado. Una vez eliminado el viernes como día del examen, el argumento se puede aplicar al jueves: como es completamente imposible que el examen tenga lugar el viernes, si en los tres primeros días de la semana aún no ha habido examen, entonces el miércoles por la noche los alumnos sabrán con certeza que va a realizarse el jueves, y dejaría de nuevo de ser inesperado. El argumento se puede repetir hasta el domingo por la noche, día en el que los alumnos estarían seguros de que el examen debe ocurrir en la mañana del lunes. La conclusión final es que las condiciones que el profesor ha impuesto al examen son contradictorias: no puede tener lugar durante la semana y ser inesperado. Los alumnos se van felices a casa y no estudian durante el fin de semana. Sin embargo, el martes de la semana siguiente el profesor, a las doce en punto, anuncia con voz grave el comienzo del examen. Los alumnos, por supuesto, no sospechaban nada y, por tanto, las premisas que anunció el profesor se han cumplido. ¿Dónde estaba entonces el error en el razonamiento de los alumnos?

La paradoja ha suscitado un gran número de artículos y una fuerte polémica con opiniones para todos los gustos: hay quien piensa que no hay paradoja en absoluto, quienes defienden que la conclusión de los alumnos contradice las propias hipótesis de su argumento, e incluso los que relacionan la paradoja con el teorema de indecibilidad de Gödel. Esta falta de acuerdo tiene su origen en la dificultad para formalizar el concepto de sorpresa, que es fundamental en la formulación de la paradoja.

Sin entrar en esta larga y vieja controversia, vamos a analizar aquí una variante que convierte el reto entre el profesor y el alumno en un juego. Supongamos que el alumno puede cada día decir si habrá o no habrá examen. Si acierta la fecha del examen recibe 8 puntos y si falla (tanto por predecir el examen en un día en que no ocurre, como lo contrario) recibe una penalización de 1 punto. Cuando el alumno dice que no hay examen un día en el que, en efecto, no lo hay, no gana ni pierde nada. Evidentemente, el objetivo del alumno es obtener la máxima puntuación mientras que el del profesor es que la puntuación sea mínima. Si el profesor pone el examen el viernes, el alumno ganará con seguridad los 8 puntos, aunque en los cuatro primeros días puede equivocarse varias veces. Por el contrario, si pone el examen el lunes, quizás el alumno no lo detecte entonces, pero los cuatro días restantes no se equivocará y no recibirá ninguna penalización. ¿Cuáles son entonces las mejores estrategias para el profesor y para el alumno?

Supongamos que sólo disponemos de dos días para poner el examen. Las posibles acciones del profesor son: poner el examen el jueves o ponerlo el viernes; mientras que el alumno puede decir que hay examen el jueves o decir que no lo hay, siendo obvia su acción del viernes. La siguiente tabla ofrece la puntuación del alumno para cada combinación de acciones:

		Acciones del profesor (día del examen)	
		Jueves	Viernes
Acciones del alumno	Examen el jueves	8	7
	No examen el jueves	–1	8

Para entender cada entrada de la tabla hay que considerar la puntuación del jueves y la del viernes. Si, por ejemplo, el alumno predice examen el jueves y el profesor lo pone el viernes, el alumno tendrá una penalización de 1 punto el jueves pero ganará 8 puntos el viernes, obteniendo un total de 7 puntos. A primera vista, parece que la acción óptima para el profesor es poner el examen el jueves, porque la columna correspondiente es la que tiene los valores más pequeños, y para el alumno decir que el examen será el jueves, puesto que la fila correspondiente es la que tiene los valores más grandes. Sin embargo, el profesor podría, siguiendo este mismo argumento, deducir que el alumno va a decir que hay examen el jueves y de ello concluir que su mejor estrategia es poner el examen el viernes. A su vez, el alumno puede pensar que el profesor piensa que él va a decir que hay examen el jueves. Puede entonces predecir que el profesor va a elegir el viernes como fecha para el examen y, con esa certeza, decir que no hay examen el jueves y ahorrarse la pérdida de un punto, alcanzando una puntuación total de 8. En teoría de juegos son comunes estas situaciones en donde un jugador piensa que el otro piensa que él piensa que el otro piensa... Son juegos en donde no hay una estrategia clara para cada jugador. Sin embargo, si se consideran combinaciones aleatorias de

estrategias, entonces el juego sí tiene una solución óptima para los dos jugadores.

Si el alumno juega la estrategia "examen el jueves" con probabilidad x y la "no examen el jueves" con probabilidad $1 - x$, mientras que el profesor pone el examen el jueves con una probabilidad y, y el viernes con una probabilidad $1 - y$, entonces, la puntuación media del alumno será:

$$p = 8xy + 7x(1 - y) - y(1 - x) + 8(1 - x)(1 - y)$$

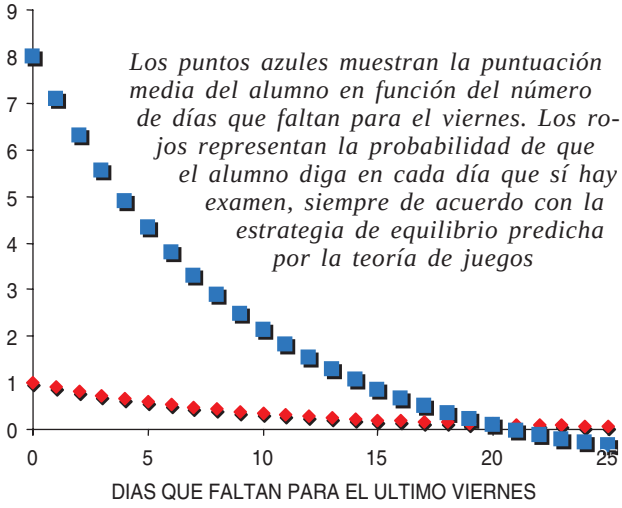
El lector puede comprobar que, si $x = 9/10$, la puntuación media es igual a 7,1 puntos, independientemente del valor de y. A su vez, para $y = 1/10$, la puntuación media es también igual a 7,1 puntos, independientemente del valor de x. En teoría de juegos esta combinación de estrategias se denomina *de equilibrio*, ya que, si un jugador la adopta, el otro no tiene posibilidad de cambiar la puntuación variando su estrategia. Por el contrario, si uno de los jugadores se aleja de la estrategia de equilibrio, el otro siempre tiene a su disposición una nueva estrategia con la que aumentar su ganancia. No obstante, hay que tener en cuenta que estas estrategias combinadas cobran sentido sólo cuando el juego se repite un gran número de veces: un 10 % de ellas, el profesor pondrá el examen el jueves y el 90 % restante el viernes. El alumno, por su parte, dirá un 90 % de las veces que el examen es el jueves y un 10 % dirá que no hay examen el jueves. De este modo ambos se aseguran una puntuación media de 7,1 en cada juego. Sin embargo, es erróneo deducir de ello que, en una situación que consistiera en una sola ronda del juego, el profesor deba poner el examen el viernes y el alumno proclamar que lo ha puesto el jueves. Con ello volveríamos al argumento circular anterior. La estrategia de equilibrio es válida sólo si las acciones se eligen al azar, con las probabilidades prescritas.

En cualquier caso, es curioso que en la estrategia de equilibrio el profesor elija el viernes un 90 % de las veces. La razón es que el alumno está fuertemente inclinado a decir que hay examen el jueves para evitar la puntuación –1. La estrategia de equilibrio presupone esta fuerte inclinación y saca partido de ella poniendo el examen el viernes un gran número de veces, para así evitar que el alumno alcance la puntuación máxima de 8.

Supongamos que ahora disponemos de tres días para el examen. ¿Qué ocurre el miércoles? La tabla de puntuación sería ahora la siguiente:

		Acciones del profesor el miércoles	
		Examen	No examen
Acciones del alumno el miércoles	Examen	8	6,1
	No examen	–1	7,1

Para construir la tabla hemos tenido en cuenta que, si el profesor no pone el examen el miércoles, el jueves tienen que jugar de nuevo a un juego cuya ganancia media hemos ya demostrado que es 7,1 puntos.



Los puntos azules muestran la puntuación media del alumno en función del número de días que faltan para el viernes. Los rojos representan la probabilidad de que el alumno diga en cada día que sí hay examen, siempre de acuerdo con la estrategia de equilibrio predicha por la teoría de juegos

Por tanto, si el alumno dice que sí hay examen el miércoles, pero el profesor no lo pone en este día, entonces tendrá una penalización de un punto y, al día siguiente y con el juego del jueves, ganará, en media y si sigue su estrategia de equilibrio, 7,1 puntos. Su puntuación total media es entonces de 6,1 puntos. Las estrategias de equilibrio en este nuevo juego son $x = 81/100$, $y = 1/10$, y la puntuación media con estas estrategias es ahora de 6,29 puntos. La estrategia del profesor no cambia con los días: siempre decide un 10 % de las veces poner el examen y un 90 % no ponerlo (salvo el viernes, día en el que tiene que ponerlo forzosamente si aún no lo ha hecho). La estrategia del alumno sí cambia a lo largo de la semana: al principio debe decir que hay examen con menos probabilidad que al final. En la figura se puede ver cómo cambia la puntuación media y la estrategia del alumno en función de los días que le separan del viernes. He prolongado la gráfica más allá de los cinco días de la semana para que pueda verse cómo, si el número de días en los que el examen puede ocurrir llega hasta 22, la puntuación media del estudiante se hace negativa.

Se pueden pensar otras variantes del juego: cambiando las puntuaciones, permitiendo que el alumno pueda decir si hay examen o no sólo una o dos veces a lo largo de la semana, etc. Los lectores con conocimientos de teoría de juegos pueden investigar sobre ello y enviarme sus conclusiones. Pero, ¿aporta algo el juego que hemos discutido aquí, o alguna de sus variantes, al análisis de la paradoja original? El juego muestra algo que ya era evidente: que existe la posibilidad de que los alumnos yerren diciendo que no hay examen un día en que lo hay, lo cual es una prueba de la existencia de un examen realmente inesperado. Sin embargo, hay una diferencia sustancial entre el juego y la paradoja. En el juego se permite que el profesor ponga el examen el viernes. Se le penaliza dando 8 puntos al alumno, pero esta penalización puede compensarse con los puntos que éste pierde durante el resto de la semana. Aun así, el juego da una idea de cómo la puntuación media y, por tanto, la capacidad de acertar de los alumnos, va disminuyendo cuando nos alejamos del último viernes.

JUEGOS MATEMÁTICOS

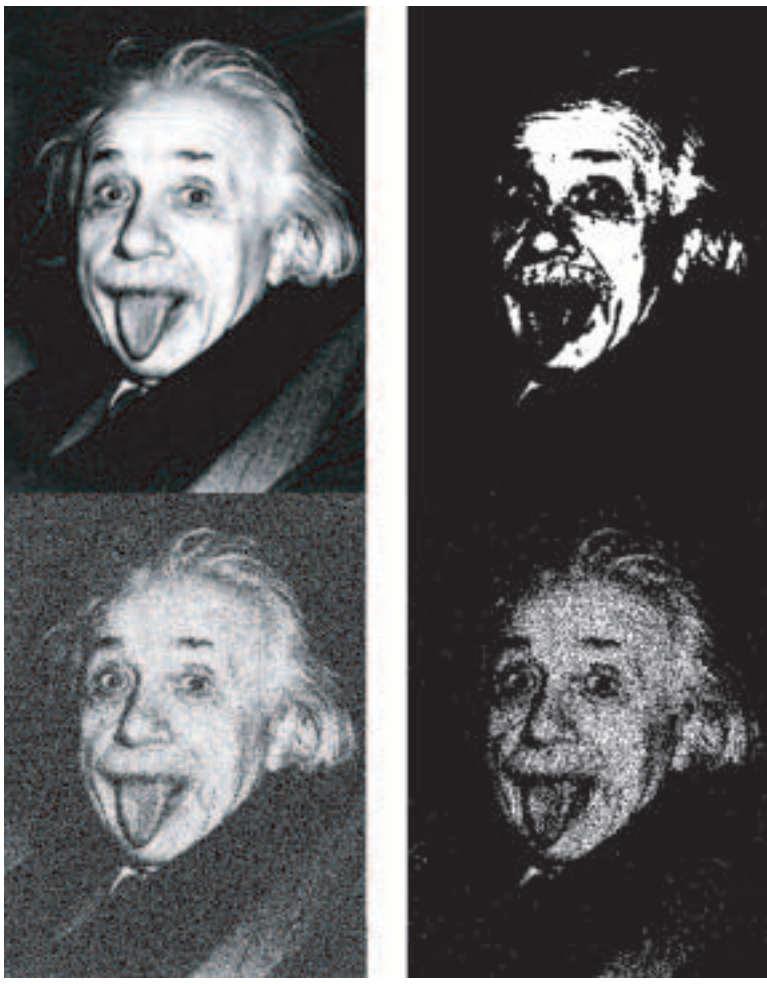
Juan M.R. Parrondo

Ruidos reveladores

Entender a alguien que nos habla en una estación en el momento en que llega un tren es una tarea difícil. El estruendo del tren se suma al sonido de la voz de nuestro interlocutor y es imposible separar la "señal" del "ruido". En cualquier canal de comunicación físico, desde las ondas sonoras con las que nos hablamos los unos a los otros hasta las ondas electromagnéticas en las que viajan las imágenes que vemos en la televisión, existen fuentes inevitables de interferencias o "ruido" que se añade a la señal transportada. Este ruido dificulta la recepción del mensaje y puede incluso distorsionarlo y hacerlo ininteligible. En la tecnología de comunicaciones, el ruido es algo poco deseable y se han desarrollado métodos para contrarrestarlo, como la teoría de corrección de errores. Sin embargo, existen situaciones en las que, sorprendentemente, el ruido puede favorecer la escucha.

Supongamos una cierta "retina artificial" que, al contemplar una imagen, pueda únicamente transmitir por cada punto un bit de información, es decir, puede tan sólo transmitir si el punto es negro o blanco. Supondremos que esta retina, al recibir la luz de una imagen en blanco y negro pero que contiene toda la escala de grises, decide que cada punto de la misma es negro si el gris correspondiente a ese punto está por debajo de un cierto umbral y blanco si está por encima. Este proceso se puede realizar en la mayoría de los programas informáticos de tratamiento de imágenes y fotografías y se denomina posterizar, puesto que ésta fue la técnica utilizada para obtener el famoso póster del Che Guevara en los años setenta.

Si el umbral es muy alto, la retina transmitirá una imagen prácticamente negra y será de poca utilidad. Esto es lo que ocurre con la imagen de Einstein en la figura 1. Podemos decir que la retina, con este umbral, es prácticamente "ciega". Su sensibilidad es demasiado baja y no percibe los grises que definen la figura de la fotografía. Veamos que ocurre si añadimos un poco de ruido a la imagen, es decir, si a la luminosidad de cada uno de sus puntos le añadimos un número aleatorio. Este ruido es muy parecido al que distorsiona la imagen de un canal de televisión mal sintonizado, como pueden comprobar en la figura. El rostro de Einstein se ve también distorsionado y su visión es incómoda para un ojo normal. Sin embargo, para nuestra torpe retina el ruido es bastante útil. En la posterización de la imagen ruidosa, Einstein es ya reconocible, algo que no ocurría con la posterización de la imagen nítida y sin ruido. En este ejemplo, el ruido ayuda a ver. Por supuesto, si añadiéramos una mayor cantidad de ruido, la información se perdería y la posterización no sería capaz de recuperar la imagen original. Existe una cantidad de ruido óptima, para la cual la imagen se recupera con mayor fidelidad. Este fenómeno se denomina resonancia estocástica. Ocurre en algunos sistemas que responden ante ciertos estímulos o señales externas y en los que algo de ruido añadido a la señal es capaz de mejorar la respuesta.



1. Einstein y sus correspondientes posterizaciones: en la columna de la izquierda se muestran las imágenes antes de posterizar y en la columna de la derecha las imágenes posterizadas. La fotografía original se muestra arriba a la izquierda y debajo la imagen con un ruido de 70 unidades (la gama de grises utilizada va de 0 a 255 y se ha tomado 225 como umbral de posterización)

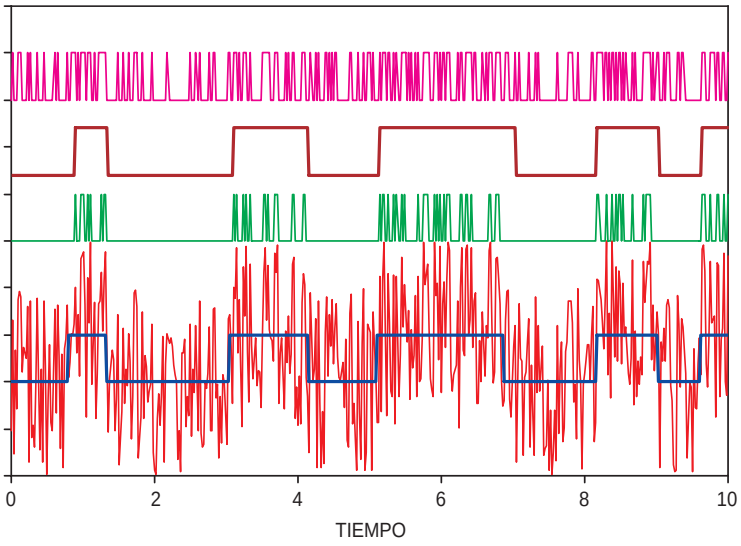
En un primer momento parecería que la resonancia estocástica es de poca utilidad. ¿Por qué debería una retina confiar en el ruido para que se desvelaran los perfiles del mundo que tiene delante? ¿No es mejor aumentar la sensibilidad, es decir, disminuir el umbral, de modo que la retina pueda ver sin necesidad de ruido?

Esta estrategia puede ser conveniente en ocasiones, pero una excesiva sensibilidad es completamente ineficaz cuando el ruido está siempre presente y es intenso. En la figura 2 se puede ver una situación de este tipo. La curva azul representa una señal que toma sólo dos posibles valores en el tiempo, 0 y 0,5, y que podría ser un mensaje en código Morse o en lenguaje binario. Como señal he elegido saltos al azar en el tiempo de una duración media de uno o dos segundos. A la señal se superpone un ruido cuya magnitud es comparable a la propia señal. Este ruido no es más que un número tomado al azar entre –1 y 1 que he sumado a la señal original. El resultado es la gráfica roja.

Si quisiéramos diseñar un sistema que detectara la señal, en principio pensaríamos en uno cuyo umbral estuviera por debajo de 0,5, es decir, un sistema que disparara si detectase una entrada por encima de 0,5. La respuesta de este sistema es la curva rosa de la figura 2. El ruido es tan intenso que este sistema está completamente confundido: dispara de un modo prácticamente independiente de la señal, ya que el ruido hace que en muchas ocasiones la entrada al sistema, es decir, la curva roja, supere el umbral aun siendo nula la señal. Sin embargo, un sistema con un umbral más alto, es decir, con una sensibilidad menor, se comporta bastante mejor. La curva verde representa la respuesta de un sistema cuyo umbral es 1. Este sistema sí capta los intervalos de tiempo en donde la señal es nula y aquellos en donde la señal es 0,5. Aún mejor es el sistema cuya respuesta viene dada por la curva marrón. En este caso hemos programado al sistema para que dispare cuando la entrada supera un umbral igual a 1 pero, una vez disparado, deja de hacerlo sólo si la entrada está por debajo de –0,5. Este sistema capta prácticamente a la perfección la señal original. ¡Pero observen que lo hace teniendo únicamente como entrada la curva roja, en la cual la señal parece completamente enmascarada por el ruido! Por supuesto, tanto el sistema de la curva verde como el de la curva marrón serían completamente "ciegos" a la señal azul si a ésta no se le añadiera ningún ruido. Estos dos sistemas de detección presentan resonancia estocástica. Son completamente inútiles para señales con poco o con mucho ruido y sólo alcanzan un régimen de detección eficaz para ciertos valores intermedios de la intensidad del ruido (aunque en este caso, bastante altos, al menos comparados con la señal).

La resonancia estocástica parece ser útil para sistemas que tienen que detectar señales muy débiles, comparables al ruido que las rodea. ¿Existe alguna aplicación práctica del fenómeno? Ha habido algún intento en ingeniería, pero sin demasiado éxito. Sin embargo, los seres vivos sí parecen utilizar la resonancia estocástica. El grupo de Frank Moss, en la universidad de Missouri, realizó experimentos con un cierto tipo de pez del Mississippi, el llamado pez raqueta. Su alimentación básica consiste en plancton formado por pequeños organismos llamados dafnias. El pez raqueta detecta las dafnias a través del campo eléctrico que generan. El río es, en lo que se refiere a campos eléctricos, un entorno bastante ruidoso y el campo eléctrico generado por las dafnias es muy débil, de modo que se trata de una situación idónea para la resonancia estocástica. Moss y sus colaboradores pudieron comprobar que el sistema de detección del pez raqueta presenta el fenómeno de resonancia. Introdujeron a un pez raqueta en un estanque en el laboratorio en el que podían generar bajo control un campo eléctrico aleatorio, es decir, un ruido, y midieron su eficacia predadora en función de la intensidad de este ruido eléctrico. El resultado fue que la eficacia del pez aumentaba al aumentar la intensidad del ruido hasta un cierto valor, a partir del cual disminuía.

Desde que se descubrió en los años ochenta, la resonancia estocástica es un campo de investigación muy intenso. El número de artículos con variaciones y nuevas propuestas de sistemas que "resuenan" con ruido o con fluctuaciones crece cada año. Y estoy seguro de que en el futuro asistiremos a nuevas y espectaculares aplicaciones del fenómeno, sobre todo en biología y en ingeniería.



2. La curva azul representa una señal distorsionada por un ruido. En rojo se puede ver la suma de la señal y el ruido, la entrada de nuestro sistema detector. Un sistema con umbral por debajo de la señal tiene la respuesta dada por la curva rosa. Sin embargo, sistemas menos sensibles responden mejor, como ocurre con la curva verde y marrón. Las escalas del eje vertical son arbitrarias y se han elegido para que se puedan ver claramente las respuestas de cada sistema

Juan M.R. Parrondo

La paradoja de Simpson

El pueblo de Calda se dividió hace tiempo en dos municipios, Calda la Vieja y Calda la Nueva. En las últimas elecciones se presentaron en ambos pueblos dos partidos, llamémoslos A y B, para no herir susceptibilidades. El partido A ganó en Calda la Vieja y el B en la Nueva, con el consiguiente disgusto de Arturo, líder comarcal del partido A, que siempre había triunfado en los dos pueblos. Nuestro líder se dispuso a analizar los escrutinios y a comparar los de los dos pueblos, para tratar de encontrar en qué sectores se habían perdido o ganado votos y las razones de la derrota en Calda la Nueva. La comparación parecía en principio sencilla, porque en los dos pueblos habían votado exactamente mil ciudadanos. Sin embargo, al político le esperaba alguna que otra sorpresa.

Arturo comenzó estudiando el voto masculino y el femenino. En Calda la Nueva, su partido había obtenido un 30 % del voto femenino y un 80 % en del masculino. No estaba nada mal, a pesar de haber perdido. La sorpresa llegó cuando quiso comparar estos resultados con los del pueblo donde había ganado, Calda la Vieja. Allí sólo había obtenido el 20 % del voto femenino y el 75 % del voto masculino. En los dos sectores el porcentaje de votos en Calda la Nueva era superior al de Calda la Vieja. Sin embargo, en este último se había ganado y en el primero se había perdido. ¿Cómo es esto posible?

Nuestro líder político está siendo víctima de la llamada *paradoja de Simpson*, un fenómeno bastante conocido en estadística. Veamos, a partir de los votos obtenidos por el partido A en cada pueblo, cómo es posible que ocurra esta paradoja. En la siguiente tabla se muestran los votos de cada partido:

Partido A / Partido B	Mujeres	Hombres	Total
Calda la Nueva	210/490	240/60	450/550
Calda la Vieja	80/320	450/150	530/470

Como ya hemos dicho, el total de votos en cada pueblo es 1000. Arturo había calculado bien los porcentajes. En Calda la Nueva su partido había obtenido 210 de los 700 votos femeninos, es decir, un 30 %, y 240 de los 300 votos masculinos, es decir, un 80 %. A su vez, en Calda la Vieja habían obtenido 80 de 400 votos femeninos, es decir, un 20 %, y 450 de los 600 votos emitidos por hombres, es decir, un 75 %. La siguiente tabla resume estos porcentajes, que se pueden obtener inmediatamente de la tabla anterior:

Porcentaje de votos para el partido A	Mujeres	Hombres	Total
Calda la Nueva	30%	80%	45%
Calda la Vieja	20%	75%	53%

En esta tabla se encuentra la esencia de la paradoja. Los porcentajes por sectores son superiores en Calda la Nueva y, sin embargo, el total es inferior. Parece imposible que pueda ocurrir algo así, pero estos números son los que se obtienen del reparto de votos indicado en la primera tabla.

La razón de este extraño comportamiento estriba en que se están comparando porcentajes parciales de sectores que no tienen el mismo tamaño. Si escribimos una tabla con la participación en cada pueblo y en cada sector:

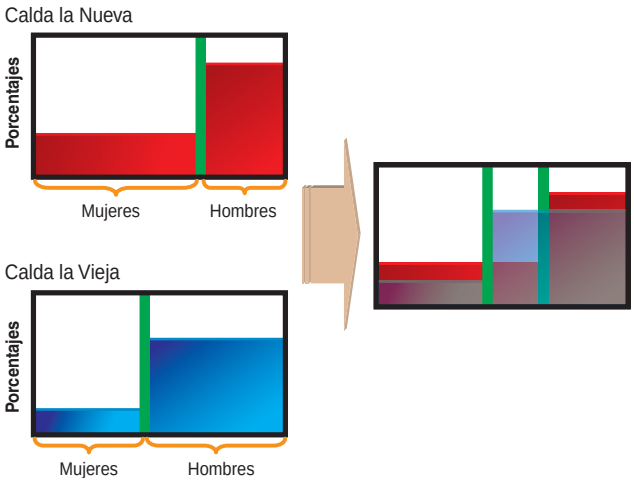
Participación	Mujeres	Hombres	Total
Calda la Nueva	700	300	1000
Calda la Vieja	400	600	1000

vemos que el voto femenino en Calda la Nueva ha sido muy superior al masculino, lo que ha hecho prevalecer su opción mayoritaria, que era el partido B. Por el contrario, en Calda la Vieja son los hombres los que han depositado más votos en la urna y, como lo han hecho en un alto porcentaje al partido B, ha sido éste el ganador.

La solución de la paradoja se hace aún más evidente en la representación gráfica de la figura 1. Los rectángulos representan el total de votos emitidos en cada pueblo. Separamos los femeninos de los masculinos con una línea verde vertical, y los que han sido para el partido A con líneas horizontales en cada sector (*rojas en Calda la Nueva y azules en la Vieja*). El porcentaje o fracción de votos en cada sector y en cada pueblo viene dado por la altura que alcanza la línea horizontal correspondiente, mientras que el porcentaje total en cada pueblo viene dado por el área coloreada.

La versión geométrica de la paradoja es que las líneas horizontales rojas pueden estar por encima de las azules y, sin embargo, el área roja puede ser *menor* que el área azul. Esto es perfectamente posible, como podemos ver en el tercer rectángulo que no es más que la superposición de los dos anteriores: el área azul posee un rectángulo entre las dos líneas verdes que puede compensar el exceso de área roja en los extre-

Representación gráfica de la paradoja de Simpson. Las líneas horizontales rojas están por encima de las azules, es decir, en cada sector, los porcentajes de votos al partido A son mayores en Calda la Nueva. Sin embargo, el área total azul es mayor que la roja, lo que equivale a que el porcentaje total de votos al partido A sea mayor en Calda la Vieja



mos. La representación gráfica ayuda también a encontrar conjuntos de datos en los que se dé la paradoja. Para ello el área del rectángulo azul tiene que ser grande y, por tanto, lo debe ser su base, que es la distancia entra las dos líneas verdes y su altura. Es decir, tienen que ser grandes tanto la diferencia entre los tamaños de los sectores en un pueblo y otro, como la diferencia entre los porcentajes de cada sector.

La paradoja de Simpson no es sólo una curiosidad matemática. Puede darse en situaciones reales, como en la evaluación de terapias y fármacos, y nos advierte de que hay que interpretar con cuidado los datos estadísticos. El caso más conocido ocurrió en la Universidad de California en Berkeley. Para analizar si había algún tipo de discriminación de género, se realizó un estudio sobre el porcentaje de solicitudes de admisión rechazadas en los distintos departamentos. Los datos globales de la universidad indicaban que las mujeres eran rechazadas más frecuentemente que los hombres. Sin embargo, este sesgo se invertía en los datos de los departamentos, en donde la frecuencia de rechazo de los hombres era mayor. Se trataba de un nuevo caso de la paradoja de Simpson. La respuesta se encontraba en que hombres y mujeres no solicitaban la admisión en los departamentos de modo uniforme, sino que ellas lo hacían en mayor medida en los departamentos en donde era más difícil ser admitido.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Un mundo sin números

"El libro de la Naturaleza está escrito en caracteres matemáticos". Así formulaba Galileo el principio fundacional de la física contemporánea, que se ha extendido más tarde a la ciencia en general y hasta a la vida cotidiana. Hoy da la impresión de que el único conocimiento válido y objetivo es el cuantificable. El debate político parece errático y vago si no se apoya en cifras económicas y sociológicas. Para explicarnos los efectos de una operación o de una terapia, el médico utiliza la probabilidad y la estadística. La máxima de Galileo tiene más vigencia que nunca: los números y las matemáticas se encuentran por doquier y están ahí para describir y explicar el mundo.

Sin embargo, la mayoría sigue teniendo grandes dificultades para aprender matemáticas. Esa dificultad no proviene sobre todo de una falta de aptitudes. Niños con problemas en matemáticas son capaces de asimilar y utilizar reglas de algunos juegos más complejas que las que rigen el cálculo elemental. En la mayoría de los casos, la dificultad proviene más bien de un cierto estupor que se apodera de nosotros en nuestro primer contacto con las matemáticas. El estupor de no entender qué tienen que ver los números con el mundo que nos rodea. Las matemáticas no son un juego, sirven para algo, y, sin embargo, el niño las percibe, y con razón, como algo completamente ajeno a su mundo, en flagrante contradicción con la máxima de Galileo.

Las matemáticas, el pensamiento matemático, ejercen una "violencia" sobre el pensamiento del niño, o de cualquiera que se acerque a ellas. Un ejemplo ilustrativo son los números negativos, la primera gran abstracción de la matemática, o, en palabras de Klein, "el paso de las matemáticas concretas a las formales". Para enseñar los números negativos a un niño, podríamos diseñar un juego en donde un dado hiciese avanzar una ficha y otro la hiciera retroceder; le explicaríamos que es lo mismo "retroceder 5 casillas" que "avanzar –5 casillas". El problema es que el niño puede entender esto como un juego de palabras absurdo, una complicación innecesaria, y comience a crearse en él la idea de que existe una distancia insalvable entre la vida y las matemáticas.

¿Tiene razón nuestro alumno suspicaz? ¿Para qué decir "avanzo –5" en lugar de "retrocedo 5"? Nosotros sabemos que la verdadera potencia de los números negativos radica en que nos permiten describir el movimiento de la ficha con un sólo verbo: "avanzar". Sin embargo, para el niño la vida está en las palabras, no en los números. Quitar palabras para dejar espacio a los números o a los símbolos es un acto de violencia intelectual, una andanada contra nuestro sistema cognitivo más básico y natural. La educación matemática de hoy en día no tiene en cuenta muchas veces esta violencia; está obsesionada con el aprendizaje de conceptos formales a una edad demasiado temprana, cuando los niños no pueden entender su potencia y utilidad, y es por eso que muchos acaban teniendo la sensación de que las matemáticas "no son lo suyo".

Si Galileo estuviera en lo cierto, si las matemáticas fueran el lenguaje necesario para entender el mundo, todos haríamos un esfuerzo considerable por aprenderlas. Si usted va a pasar el resto de su vida en China, dedicaría gran parte de su tiempo y de su energía a aprender chino, por muy difícil que sea o por muy mal que se le den los idiomas. Además, en poco tiempo poseería ciertos rudimentos de chino, puesto que en su vida diaria tendría que aplicar constantemente lo aprendido. La utilidad de lo estudiado en el aula sería evidente en cada momento de su vida cotidiana. Sin embargo, con las matemáticas no ocurre lo mismo. Ocurre, de hecho, todo lo contrario. Sobre todo para un niño, las matemáticas son algo propio de la clase, algo restringido al espacio delimitado por las paredes del aula, que pierde todo sentido fuera de ellas y carece de utilidad para la comprensión de su realidad más inmediata.

Y así fue en Occidente durante muchos siglos. En la Edad Media y en el mundo clásico, los números tenían poco que ver con la realidad. Se vivía en un mundo percibido y entendido de forma cualitativa. Algo que cambió drásticamente en un período de tiempo relativamente corto, que va desde 1250 hasta 1350. En ese breve intervalo de tiempo, se desarrolló más la polifonía y se inventaron el reloj mecánico, el cañón y, probablemente, la perspectiva y los libros de contabilidad. Inventos que ofrecían una imagen cuantitativa de la realidad o que obligaban a crearla y manejarla. Antes de ellos, tiempo y espacio eran esencialmente cualitativos. En un libro delicioso, *La medida de la realidad,* el historiador Alfred W. Crosby describe este mundo cualitativo medieval y la transición a lo que él denomina *pantometría*, la obsesión por medir, por convertir en número cualquier aspecto de la realidad.

El hombre medieval percibía el tiempo de un modo muy diferente a nosotros. El tiempo histórico no se basaba en cronologías exactas, sino que sólo estaba jalonado por grandes acontecimientos, como el diluvio universal o la vida de Jesucristo. Era más "un escenario" que una línea de acontecimientos. El tiempo cotidiano era también bastante cualitativo. Existían las horas, pero su duración distaba mucho de ser exacta. Por ejemplo, había siempre doce horas de día y doce de noche, tanto en verano como en invierno, con lo que las horas diurnas veraniegas eran mucho más largas que las del invierno. A su vez, lo único que marcaba ciertos intervalos de tiempo de forma pública eran las campanas de los monasterios, que indicaban las "siete horas canónicas": maitines, prima, tercia, sexta, nona, vísperas y completas. La poca precisión con la que se tocaban estas horas canónicas queda viva-

mente demostrada por la etimología de la palabra inglesa *noon*, que hoy en día denota las doce del mediodía, pero cuyo origen se encuentra en una de esas horas: la *nona*. La nona correspondía a la novena hora a partir del amanecer, es decir, más o menos a las tres de la tarde. Sin embargo, con el tiempo la llamada a nonas se fue adelantando, debido probablemente al hambre de los monjes, ya que era ésa la hora en la que podían tomar el primer bocado del día. Los períodos de tiempo menores que una hora, como la duración de la cocción en una receta, se medían mediante fórmulas cualitativas que han pervivido hasta nuestros días, como el rezo de un padrenuestro para un huevo pasado por agua. No había minutos ni segundos, y las horas se hinchaban y contraían dependiendo de la época del año y de las actividades de los monjes.

El espacio medieval era también muy distinto del espacio moderno. Los mapas estaban llenos de lo que ahora calificaríamos como distorsiones. Sin embargo, estas distorsiones no siempre se debían a una falta de pericia o de medios técnicos, sino que eran consecuencia lógica de la función misma del mapa en la Edad Media: no querían guiar guiar al viajero o al navegante, sino dar una imagen del mundo, "facilitar información sobre lo que estaba cerca y lo que estaba lejos, y lo que era importante y lo que no lo era". Los mapas de ayuda a la navegación, como los primeros mapas *portolanos*, en donde se detallaban con precisión los puertos a lo largo de las costas europeas y norteafricanas, surgieron a principios del siglo XIII. Antes de esa fecha, los mapas apenas si se utilizaban con ese fin.

En un mundo no medido, los números tenían poca cabida y, cuando se utilizaban, cobraba más importancia su simbología que su modesta referencia a la simple cantidad. Veamos por ejemplo cómo describe algunos números Bartolomé Anglicus, en el siglo XIII, traducido por Fray Vicente de Burgos al español en 1494:

"El numero de ocho es conplido por una unidad puesta sobre siete y es compuesto de dos vezes quatro que son numeros pares y de V y tres que son nones y de VII y uno y significa la abundançia de gloria que avran los que avran las siete virtudes o los que avran los siete dones del Espiritu Santo.

"Una unidad puesta sobre VIII haze IX que es numero compuesto de tres vezes tres y significa el estado y la gloria de las tres hierarchias de los angeles de las quales cada una ha conformidad a la Trinidad gloriosa y son permanesçientes en la suma deidad sin medio alguno.

"El numero de X añade una unidad sobre IX y es el fin y termino de todos los numeros simples ca quien pasa X contando torna a uno y dende a dos y assi de los otros. El numero de X que es el fin de todos numeros sinples es el comienço de todos numeros compuestos y significa Dios que es fin y comienço de todas creaturas sean sinples como los angeles o compuestas como los honbres."

La notación es en parte romana, una notación que de por sí no facilita los cálculos. El número no sólo estaba alejado de la vida cotidiana, sino que era un artefacto místico, relacionado con los espacios sagrados y con el arte más abstracto, la música. De ahí que los cálculos fueran reducidoslo: se podía hacerlos sin apenas salirse del complejo sistema de numeración romana, o con los dedos u otras partes del cuerpo y, a partir del año 1000, con el ábaco, importado de Oriente gracias a los musulmanes españoles.

Un "mapa" medieval de los llamados OT, que representa los tres continentes conocidos en la época separados por dos mares en forma de T. El mapa está orientado con el este hacia arriba, ya que, siendo la dirección de la salida del Sol, era el punto cardinal de mayor rango

Este mundo medieval cualitativo y sin números, este "modelo venerable", como lo llama Crosby, no ofrecía una imagen más grosera o más ingenua de la realidad. En palabras del propio Crosby, "mostramos desdén ante sus errores", pero nuestro verdadero problema con este modelo estriba en que sea "dramático, incluso melodramático: Dios y el Designio se ciernen sobre todo". En la actualidad "queremos (o pensamos que queremos) explicaciones de la realidad desprovistas de emoción, tan anodinas como el agua destilada". Por el contrario, "los europeos de la Edad Media y del Renacimiento, al igual que el chamán, al igual que todos nosotros parte del tiempo y algunos de nosotros todo el tiempo, querían explicaciones que fuesen concluyentes de modo inmediato y satisfactorias desde el punto de vista emocional. Anhelaban un universo que, como dice Camus, pueda amar y sufrir".

Muy probablemente, la experiencia más inmediata del mundo para la mayoría de las personas tiene que ver antes que nada con ese modelo caduco y tosco del medievo. Nos hemos acostumbrado al tiempo y al espacio compartimentados y medidos, a cuantificar con dinero nuestros deseos y el esfuerzo que estamos dispuestos a realizar para conseguirlos. Sin embargo, y por más que le pese al viejo Galileo, en lo más íntimo y profundo, seguiremos viviendo en un mundo sin números, o donde los números son un mero juego o una útil, pero limitada, herramienta.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

El problema del secador de manos

Hace poco visité el nuevo bar de mi amigo Andrés. Después de lavarme las manos en el servicio del bar, pulsé el botón del secador, uno de esos que funciona con aire caliente y disponen de un temporizador. Como quizá le haya ocurrido al lector más de una vez, el aparato se detuvo cuando tenía las manos aún húmedas. Incómodo, volví a ponerlo en marcha, las manos se secaron mucho antes de que el temporizador detuviera el secador por segunda vez y al salir continué oyendo el zumbido machacón del motor durante un buen rato desde la barra.

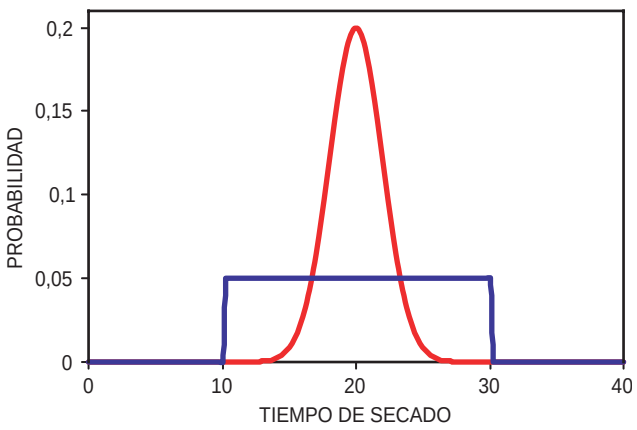
Atormentado desde hace años por esta asincronía entre los secadores y mi capacidad de frotar las manos, le grité a Andrés, mientras se seguía escuchando el murmullo del aparato a través de la puerta del servicio:

— ¿Por qué programan estos cacharros para que funcionen durante un intervalo de tiempo tan corto? Es molesto tener que volver a pulsar el botón con las manos aún húmedas, pero aún más ver cómo se derrocha energía eléctrica con el secador funcionando la mitad del tiempo para nada.

Andrés contestó sin dejar de dar brillo a las copas, como si hubiese tenido la misma conversación una decena de veces:

— No todo el mundo tarda lo mismo que tú. Hay quienes lo ponen en marcha cuando ya casi están abriendo la puerta para volver al bar. Y otros parece que hacen la colada allí mismo, del tiempo que se pasan debajo del aire caliente. Supongo que el fabricante del secador habrá tomado una especie de media. Piensa también que cuanto más corto sea el intervalo de tiempo, menos energía se derrochará.

— Veo por dónde vas. Si el intervalo fuese de un segundo, por ejemplo, en realidad no se podría nunca derrochar más energía que la gastada en un segundo.



1. Distribuciones del tiempo que necesitan los usuarios para secarse completamente las manos: uniforme (en azul) *y campana de Gauss* (en rojo)

Pero un secador programado así sería una auténtica pesadilla, tendríamos que estar pulsando constantemente el botón de arranque.

Al día siguiente volví a visitar el bar de Andrés armado con toda clase de gráficas y análisis que había realizado la noche anterior. Mi idea era encontrar la duración que debe programarse en el secador para que se derroche la mínima cantidad de energía. Como bien intuía Andrés, la solución sería siempre que el intervalo fuera lo más pequeño posible, a no ser que se tenga en cuenta que al pulsar el botón también existe una cierta pérdida de energía o, si se quiere, un desgaste en la maquinaria del aparato. Decidí que esta pérdida fuera de 0,1 céntimos de euro por pulsación. También es necesario hacer alguna suposición sobre el tiempo que tarda la gente en secarse las manos. He considerado varias posibilidades: que todos tarden lo mismo, 20 segundos, que los tiempos se distribuyan de modo uniforme entre 10 y 30 segundos, o que lo hagan siguiendo una campana de Gauss, también alrededor de 20 segundos. En todos los casos, el valor medio del tiempo es de 20 segundos. Estas dos últimas distribuciones se muestran en la figura 1. Finalmente, supondré que mantener el aparato encendido 20 segundos cuesta un 1 céntimo de euro, es decir, cada segundo nos cuesta 1/20 céntimos. Por tanto, si el secador está programado con una duración de d segundos, el gasto total cada vez que se conecta será de $0{,}1 + d/20$ céntimos de euro, 0,1 por la pulsación y $d/20$ por el funcionamiento durante d segundos.

Con todas estas suposiciones, necesarias para completar el modelo, se puede calcular el gasto total medio de cada persona que utiliza el secador en función de la duración de funcionamiento programada. El cálculo no es demasiado complicado, al menos para el caso en el que todos los usuarios tardan 20 segundos en secarse. Si un individuo tarda un tiempo T en secarse y la duración programada es d, entonces pulsará el botón un número de veces dado por la fórmula:

$$\mathrm{PE}\ (T/\mathrm{d})$$

en donde PE significa aquí “parte entera del número que está entre paréntesis”, es decir, el número entero que queda si eliminamos todos los decimales que van detrás de la coma (por ejemplo, la parte entera de 2,43 es 2). El gasto para esta persona es entonces:

$$(0{,}1 + d/29) \times \mathrm{PE}\ (T/\mathrm{d})$$

Calcular el gasto en el caso en que todos los individuos se secan las manos en 20 segundos es muy sencillo: basta hacer $T = 20$ en la fórmula anterior. El resultado es la curva verde de la figura 2. Como vemos, la duración óptima es evidentemente $d = 20$ se-

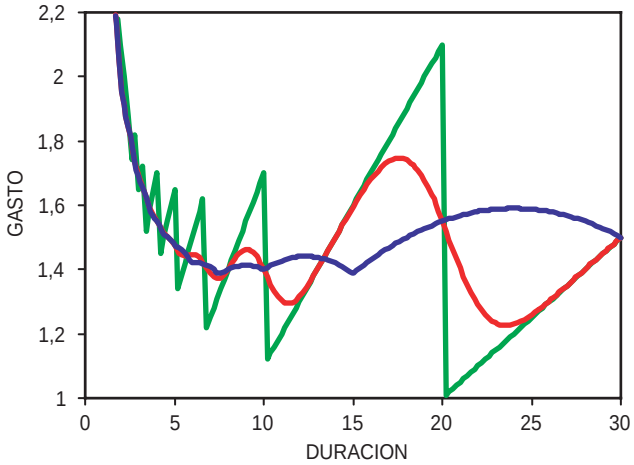
gundos, ya que entonces a todos los usuarios les basta una pulsación para secarse. En los casos en los que hay una distribución de tiempos, como las dadas en la figura 1, es necesario promediar la fórmula anterior utilizando la distribución correspondiente. El cálculo es bastante más difícil y hay que hacerlo con ordenador. La curva roja muestra el resultado para la campana de Gauss y la azul el correspondiente a la distribución uniforme. ¿Cuál es finalmente la duración óptima? Depende mucho de la distribución. De hecho, es sorprendente que la dependencia sea tan acusada. Por ejemplo, en el caso de la campana de Gauss, la duración óptima es de 23,5 segundos, con un gasto medio aproximado de 1,225 céntimos de euro por secado. Pero esa misma duración produciría un gasto de 1,59 céntimos si la distribución fuese uniforme.

Un poco desanimado por estos resultados, pero con la obstinación típica del científico, le dije a Andrés:

— Lo mejor que podemos hacer es apostarnos durante varios días en la puerta del servicio y medir cuánto tarda cada cliente en secarse. Así tendríamos la distribución real, podríamos calcular la duración óptima y programarla en el secador.

Mi amigo Andrés, que no es muy experto en matemáticas, pero tiene la sabiduría de un buen barman, respondió con tranquilidad sin dejar de sacar brillo a las copas:

— Todo esto está muy bien, pero me parece que no va a sernos muy útil. Lo más probable es que eso que llamas distribución de tiempos *dependa* de la propia duración del secador. Si a un tipo se le para el secador con las manos ligeramente húmedas, quizá maldiga al fabricante, pero lo más probable es que no lo conecte de nuevo y se vaya a su mesa sacudiéndolas para que acaben de secarse.



2. Gasto medio por individuo que utiliza el secador, en función de la duración con la que está programado. La curva verde corresponde al caso en que todos los individuos necesitan exactamente 20 segundos para secarse, la azul corresponde a la distribución de tiempos uniforme y la roja a la campana de Gauss

Definitivamente, me di por vencido. Lo que Andrés intuye es un caso de retroalimentación y un problema de ese tipo es bastante más difícil de resolver. Desde entonces no he vuelto a pulsar dos veces el botón del secador en el bar de mi amigo y salgo siempre sacudiendo las manos al aire.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La paradoja de la Biblioteca de Babel

El gran emperador de Babel ordenó a sus escribanos que caligrafiaran uno a uno los volúmenes de una grandiosa Biblioteca, una Biblioteca que contendría todos los libros de 500 páginas escritos y por escribir. Las instrucciones que dio al jefe de los escribanos eran muy sencillas. Deberían tomar las 27 letras del alfabeto, el espacio y los signos de puntuación "," "." ";" ":" "¡" "!" "¿" "?", y hacer todas las combinaciones posibles que se ajustaran a las 500 páginas de cada libro. Empezarían con un libro en blanco; después con un libro que sólo tuviera la letra "a" en su comienzo estando en blanco el resto. Otro con la letra "b", y así sucesivamente. Luego se empezaría con pares "aa", "ab", "ac", etc. El último de los volúmenes sería un libro cuyas 500 páginas consistirían en una sucesión interrumpida de signos de interrogación "?". Entre el primer libro en blanco y este último y misterioso libro lleno de interrogaciones estarían todos los libros escritos y por escribir. La mayoría serían ininteligibles o absurdos, pero en alguno de los estantes de la Biblioteca de Babel estarían *Don Quijote*, *La Crítica de la Razón Pura* o la exposición (sin fórmulas matemáticas) de la teoría de la relatividad.

El emperador estaba muy contento con su Biblioteca hasta que un joven estudiante pidió audiencia y le habló así:

"Majestad, habéis creado una gran obra con vuestra Biblioteca, puesto que encierra todas las verdades (y todas las mentiras), todas las historias y todas las bellezas del mundo que pueden expresarse con palabras. Pero hay algo en ello que no acabo de entender. Las instrucciones que disteis a vuestros escribanos para crear la Biblioteca se pueden dictar en menos de cinco minutos y cabrían en una simple hoja de papel. Esa hoja de papel contendría las instrucciones necesarias para crearla de nuevo. Si con unas instrucciones tan breves podemos crear todas las verdades (y todas las mentiras) del mundo, entonces cada una de estas verdades, o cada uno de esos libros, puede crearse con esas mismas instrucciones, o incluso con menos, puesto que cada libro es sólo una ínfima parte de la Biblioteca. Con su Biblioteca, Su Majestad ha demostrado que todas y cada una de las verdades y las historias de los hombres caben en una simple hoja de papel."

El emperador oyó al estudiante con atención, sin saber cómo responder. Tras despedirlo, se encerró durante semanas en los aposentos más retirados del palacio, sumido en una profunda confusión.

El argumento del joven estudiante parece absurdo, pero no puede descartarse fácilmente, y el Emperador lo sabía. Es cierto que las instrucciones para la creación de la Biblioteca caben en una hoja de papel. Por otro lado, como *El Quijote*, por ejemplo, es uno de los volúmenes de la Biblioteca, ¡está claro que sólo hace falta seguir esas instrucciones para escribirlo! ¿Cómo es posible?

Esta misma paradoja ocurre en un campo de las matemáticas, la *complejidad algorítmica*, creada por los matemáticos Kolmogorov, Chaitin y Solomonoff. Ya hablamos de ella cuando analizamos la relación entre la compresión de ficheros de datos y la teoría de la información (*Juegos matemáticos* de septiembre de 2001). La complejidad algorítmica de una cadena de símbolos es la cantidad de instrucciones que hay que dar a una persona o a un ordenador para que reconstruya dicha cadena. Si la cadena es muy grande pero responde a una determinada pauta lógica, su complejidad algorítmica es pequeña. Por ejemplo, la complejidad algorítmica del número de dos mil dígitos:

27272727 ... 27

es muy pequeña. Porque bastan las instrucciones "escriba 27 mil veces seguidas" para reconstruirla. Sin embargo, el número:

1222112121221112122 ...

que se construye tirando una moneda dos mil veces y poniendo un 2 si sale cara o un 1 si sale cruz, tiene una complejidad algorítmica muy grande, ya que si los resultados del lanzamiento de la moneda son realmente aleatorios, lo más probable es que no tengan pauta alguna y las únicas instrucciones que permitan reconstruir ese número en concreto consistan en indicar al sujeto que quiera reconstruirlo todos los dígitos, uno por uno, que lo forman. En otras palabras, el número es incompresible: la información que contiene no puede reducirse a unas instrucciones más breves que el propio número; la única forma de describirlo es detallar, una por una, todas las cifras que lo forman.

La paradoja a la que se enfrenta el emperador de Babilonia, expuesta en términos de la complejidad algorítmica de Kolmogorov, Chaitin y Solomonoff sería la siguiente:

¿Cómo es posible que la complejidad algorítmica de una secuencia sea muy pequeña y que, sin embargo, la complejidad de un subconjunto de dicha secuencia sea grande? O: ¿cómo es posible que la complejidad algorítmica de la Biblioteca sea pequeña mientras que la complejidad de uno de sus volúmenes, *El Quijote* por ejemplo, es grande?

Para ver cómo y por qué es esto posible, consideremos la cadena C de dígitos formada por la conca-

tenación de todos los números entre 0 y $10^{2000} - 1$. Para construir esta cadena *C* escribimos primero dos mil ceros, luego 1999 ceros y un 1, luego 1999 ceros y un 2, y así sucesivamente. Es decir, concatenamos todos los números de dos mil cifras, empezando por el cero, el 1, el 2, etc., hasta llegar al $10^{2000} - 1$. Esta inmensa cadena contiene

$$2000 \times 10^{2000} = 2 \times 10^{2003}$$

dígitos, pero su complejidad algorítmica es muy pequeña. Ahora bien, en la cadena *C* aparecen como subcadenas todos los números entre cero y $10^{2000} - 1$, y alguno de estos números —como el aleatorio que hemos visto antes— puede tener una complejidad algorítmica mucho mayor. Nos encontramos de nuevo con la paradoja de la Biblioteca de Babel.

La solución de la paradoja es la siguiente. Piensen en el número aleatorio de dos mil cifras que hemos dado antes. Este número —llamémoslo *n*— está dentro de la cadena *C*. Podríamos utilizar la pequeña complejidad algorítmica de *C* para describir *n*. Las instrucciones para reconstruir *n* serían: (*a*) constrúyase *C* y (*b*) encuéntrese *n* dentro de *C*. Sabemos que las instrucciones para llevar a cabo el paso (*a*) son muy breves. Pero, ¿qué ocurre con el paso *(b)*? Dentro de la cadena *C*, el número *n* empieza en el sitio:

$$2000 \times n$$

Luego, para precisar con absoluta exactitud dónde empieza *n*, ¡tendríamos que describir con absoluta exactitud el propio número *n*! Con lo cual utilizar la cadena *C* es completamente inútil: no nos ahorra ninguna instrucción.

El emperador de Babilonia encontró esta solución a la paradoja, pero no por sí solo, reflexionando en su palacio. La encontró gracias a uno de sus bibliotecarios. Después de muchos días de aislamiento y de tristes pensamientos acerca de las palabras del estudiante, decidió consultar por primera vez algún volumen de la Biblioteca (no lo había hecho nunca hasta entonces). Se acercó a la gran puerta dorada y preguntó al bibliotecario:

—He oído hablar de un libro divertido y a la vez revelador. Lo escribirá un tal Cervantes dentro de unos siglos, y trata de las aventuras de un caballero andante.

El bibliotecario le respondió sonriente:

—Sí, Majestad. Conozco ese libro. El escribano que lo transcribió me habló muy bien de él y recuerdo perfectamente en qué pasillo de la Biblioteca se encuentra. Os indicaré cómo llegar hasta él. Tomad este pasillo central hasta que veáis un gran cartel con la letra "e". Allí deberéis girar a la derecha y caminar un trecho no muy largo, hasta llegar al pasillo presidido por un cartel con la letra "n". Caminad sólo unos metros, porque enseguida deberéis tomar a la izquierda el pasillo correspondiente al espacio en blanco. Luego el pasillo de la letra "u", después el de la "n", a continuación el del espacio de nuevo. Más tarde el de la "l", el pasillo de la "u", luego el de la "g", el de la "a" y el de la "r"...

Un sencillo programa de ordenador podría generar una biblioteca que incluyese todos los libros posibles. Pero en el listado de ese programa no podríamos leer El Quijote. *Ilustración tomada del CD-ROM* La Biblioteca Total, Viaje por el Universo *de Jorge Luis Borges, ©1996, Nicolás Helft & Weber-Ferro S.R.L.*

El bibliotecario se pasó semanas indicando al Emperador el sitio exacto donde se encontraba el libro. Cuando terminó su explicación, el Emperador sonrió. No necesitaba recorrer la tupida red de corredores que formaban la Biblioteca. Sentado delante del mostrador del bibliotecario, había conocido con todo detalle las andanzas de Don Quijote y Sancho y, al mismo tiempo, había encontrado la solución a la paradoja.

Es cierto que los libros de la Biblioteca se podrían ordenar de modo tal que la tarea de describir el lugar en donde se encuentra *El Quijote* no equivaliera a recitarlo letra por letra. En *El Quijote*, como en cualquier otro libro escrito en una lengua real, hay un cierto grado de redundancia. Por ejemplo, detrás de una "q" siempre aparecerá una "u", con lo cual el bibliotecario se podría ahorrar parte de las indicaciones. Redundancia significa que en la cadena de letras que forman *El Quijote* cada letra nueva no nos "pilla completamente de sorpresa" sino que es parcial o totalmente predecible, como una "u" detrás de una "q", o como la "i" y la "a" detrás de "redundanc". Esta predecibilidad o redundancia hace que el lenguaje no sea óptimo, en el sentido de expresar la mayor cantidad de información por letra. De hecho, los compresores de ficheros que utilizamos en nuestros ordenadores detectan esta redundancia para reducir el tamaño del fichero que ocupa un texto hasta en un 70 %, como ya vimos la sección de *Juegos Matemáticos* de septiembre de 2001. Sin embargo, sería probablemente imposible aprender un lenguaje que estuviera completamente exento de redundancia.

En cualquier caso, el emperador puede dormir tranquilo: la escasa complejidad algorítmica de la Biblioteca no disminuye un ápice el valor de todas las joyas que se esconden en ella, puesto que la aventura de encontrarlas es idéntica a la aventura de escribirlas.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Zenón y los camellos

Muchos lectores quizá conozcan este viejo acertijo: Un anciano dejó antes de morir los 11 camellos de su rebaño a sus tres hijos, indicando que el primero debería recibir la mitad del rebaño, el segundo una cuarta parte y el tercero una sexta parte. El albacea no sabía muy bien cómo satisfacer los deseos del difunto sin descuartizar algún camello, puesto que al dividir 11 entre dos, cuatro o seis no se obtiene ningún número entero. Pasó toda la noche reflexionando sobre el problema y finalmente encontró una solución extraña, pero incontestable.

Pidió a un amigo un camello y lo unió al rebaño. Con este rebaño de 12 camellos reunió a los tres hijos y les leyó solemnemente el testamento de su padre. Al primero le dio la mitad del rebaño aumentado: 6 camellos. Al segundo hijo le entregó un cuarto, es decir, 3 camellos. Finalmente, al último le dio un sexto del rebaño, es decir, 2 camellos. El número total de camellos entregados a los hijos es 11. Por lo tanto, después del reparto, pudo devolver el camello prestado y satisfacer los deseos del difunto.

Este modo de repartir los 11 camellos sorprende a primera vista. El "truco" del reparto estriba en que las fracciones que el anciano padre había asignado para cada hijo no suman 1, sino

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{6}=\frac{6+3+2}{12}=\frac{11}{12}$$

La suma nos indica con absoluta claridad que, para repartir 12 camellos conforme a las fracciones asignadas, sólo es necesario entregar 11.

Hasta aquí el problema clásico. Sin embargo, hace unos días, Paco Blanco, de la Universidad Complutense, me proponía una sencilla variante para solucionar el reparto íntimamente relacionada con las antiguas paradojas de Zenón. ¿Qué ocurriría si el albacea siguiera al pie de la letra las instrucciones del difunto sin hacer uso del camello adicional? Daría al primero de los hijos 11/2 de camello, al segundo 11/4 y al tercero 11/6. No sin antes malograr la integridad física de algunos de los pobres animales, el albacea se encontraría al final del reparto con 11/12 de camello sobrante, ya que el total repartido es $(11\times11)/12$. Como es un hombre honrado que desea que la herencia se entregue íntegramente a los hijos, decide repartir lo que sobra ateniéndose de nuevo a las fracciones estipuladas por el padre. Por lo tanto, da al primero de los hijos $1/2\times11/12 = 11/24$ de camello, al segundo $(1/4\times11/12)/11/48$, y al tercero $(1/6\times11/12/12 = 11/72$. Pero, después de este segundo reparto, sigue sobrándole una pequeña fracción: $11/12\times1/12 = 11/144$ de camello. Ni corto ni perezoso, vuelve a repartir este sobrante siguiendo las instrucciones del padre. Este proceso es infinito, ya que, por muchas veces que se realice el reparto de la parte sobrante, siempre quedará una pequeña fracción. En concreto, después de n repartos, al albacea le quedarán $11/12^n$ de camello.

Tras infinitos repartos, ¿cuánto habrá recibido cada hijo? La respuesta es que este reparto, aunque no lo parezca a primera vista, conduce al mismo resultado que había conseguido el albacea con la ayuda del camello adicional. Para convencernos de ello, pensemos que en todos y cada uno de los infinitos repartos, el segundo hijo *siempre* obtiene la mitad de lo que se le da el primero y el tercero *siempre* obtiene un tercio de lo que consigue el primero. Por tanto, si el primero consigue una cantidad total de camellos x, el segundo obtendrá $x/2$ y el tercero $x/3$. Como, después de los infinitos repartos, se habrán agotado los 11 camellos, la suma de lo obtenido por cada hijo debe ser igual a 11, es decir:

$$x+\frac{x}{2}+\frac{x}{3}=\frac{11x}{6}=11$$

con lo que $x = 6$ y el segundo y tercer hijo reciben, respectivamente, 3 y 2 camellos, tal y como ocurría con el reparto que hacía uso del camello adicional.

El problema está así resuelto. Sin embargo, podríamos haberlo atacado de otro modo, tratando de sumar lo que recibe cada hijo en los infinitos repartos. Por ejemplo, el primer hijo recibe una cantidad de camellos que puede escribirse como una suma infinita:

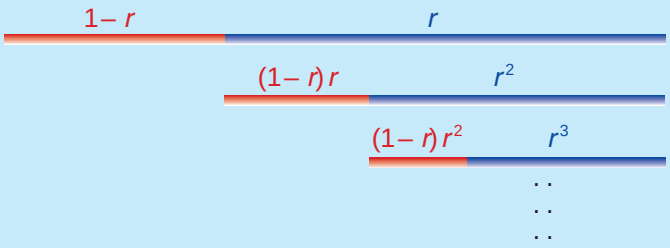
$$\frac{1}{2}\times11+\frac{1}{2}\times\frac{11}{12}+\frac{1}{2}\times\frac{11}{12^2}+\frac{1}{2}\times\frac{11}{12^3}+\cdots$$

en donde los puntos suspensivos indican infinitos términos que siguen la misma pauta. Con el argumento de las proporciones, hemos encontrado que esta suma infinita es igual a 6. Por lo tanto, hemos sido capaces de realizar esta suma de infinitos términos. La suma se puede simplificar un poco y escribirse en la forma:

$$1+\frac{1}{12}+\frac{1}{12^2}+\frac{1}{12^3}+\cdots=\frac{6\times2}{11}=\frac{12}{11}$$

Que una suma infinita pueda dar como resultado un número finito parece a primera vista extraño. Así se lo pareció hace más de dos mil años a Zenón de Elea, quien expuso esta extrañeza en forma de paradojas para evidenciar los problemas del movimiento de los cuerpos. Una de las paradojas de Zenón pone

El argumento sobre la paradoja de Zenón se puede generalizar a cualquier fracción r mediante una modificación de la idea de los repartos infinitos en donde sólo hay una persona beneficiaria del reparto. Supongamos un segmento de longitud unidad. Lo dividimos en dos partes, una de longitud $1 - r$ y otra de longitud r, y damos a la beneficiaria la primera de las dos partes (*en rojo en la figura*). Volvemos a dividir el segmento sobrante en dos fracciones dadas por $1 - r$ y r y entregamos la primera. Repetimos la división infinitas veces, tal y como se muestra en la figura, dando siempre el intervalo rojo y dejando el azul para el reparto siguiente. Después de infinitos repartos, habremos entregado todo el intervalo de longitud unidad, que debe ser igual a la suma de todos lo segmentos rojos. Es decir:



$$(1-r)(1 + r^2 + r^3 + \cdots) = 1$$

o también:

$$1 + r^2 + r^3 + \cdots = \frac{1}{1-r}$$

Las distintas fórmulas de sumas infinitas que aparecen en el texto no son más que casos particulares de ésta que acabamos de deducir.

en cuestión que un cuerpo pueda moverse, digamos, de un extremo a otro de una habitación. Su conocido argumento nos dice que el cuerpo tendrá primero que alcanzar el punto medio de su recorrido, después el punto medio de lo que le queda por recorrer, después el siguiente punto medio y así sucesivamente. El cuerpo tiene que alcanzar infinitos puntos o, dicho de otra manera, consumar un número infinito de pequeños recorridos. Si, por ejemplo, la longitud total del movimiento es un metro, el cuerpo tendrá que recorrer primero la mitad, 1/2, luego la mitad de la mitad, 1/4, luego 1/8, y así sucesivamente. Si tiene que hacer un número infinito de recorridos, y en cada uno de ellos invierte una cantidad finita de tiempo, entonces, siguiendo siempre el argumento de Zenón, el cuerpo tardará un tiempo infinito en realizar su movimiento. Con ello Zenón demostraba que el movimiento es imposible o, al menos, algo propio del mundo de las apariencias y que presenta inconsistencias lógicas.

Es curioso que Zenón no se diera cuenta de que una suma con infinitos términos puede ser finita. Por su propio argumento, y sin fijarnos en el tiempo que tarda el cuerpo en hacer cada recorrido sino sólo en las longitudes de éstos, no es difícil darse cuenta de que la suma de los intervalos en los que sucesivamente dividimos el recorrido tiene, por necesidad, que ser igual a la longitud total, es decir:

$$\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots = 1$$

Sin embargo, los griegos todavía tenían dificultades con la noción de un infinito que no fuera meramente potencial. Aristóteles, por ejemplo, propone una complicada distinción entre el segmento sin dividir y el segmento dividido infinitas veces; el primero era "actual", el segundo, "potencial", y dudaba de que la longitud de uno fuera igual a la del otro.

Lo que quizás estaba en cuestión, tanto en la mente de Zenón como en la de Aristóteles, era la propia naturaleza del tiempo y el espacio y la posibilidad de subdividirlos infinitamente. Otra de las paradojas de Zenón sobre el movimiento, la conocida como *de la flecha* y quizá la más sugerente de todas ellas, nos dice que en un instante dado una flecha está en un cierto lugar, ocupando el mismo espacio que ocuparía si estuviera en reposo. Considerando sólo ese instante, la flecha en movimiento no se distingue en nada de la flecha en reposo y podemos por tanto decir que en ese instante la flecha se haya en reposo. Si se haya en reposo en cada instante, la flecha se haya siempre en reposo, luego no se mueve nunca. Aunque parece un mero juego de palabras, la paradoja de la flecha se pregunta por algo que aún es misterioso para la física, al menos para la no relativista: si es posible generar un universo dinámico, en movimiento, a partir tan sólo de una sucesión continua de instantes y qué tipo de conexión causal o de continuidad hace falta para ello. De todos modos, para muchos físicos esta pregunta no tiene sentido. Más bien pensamos que el instante es una idealización, carente de realidad física. Recientemente, un joven estudiante de física australiano, Peter Lynds, ha recibido una considerable, pero en mi opinión no merecida, atención de los medios de comunicación por analizar las paradojas y proponer esta vieja solución: la negación del instante.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Cita a ciegas

Aunque al lector ajeno al mundo de la investigación le cueste creerlo, una de las partes más importantes —y más leídas— de un artículo científico es la llamada lista de referencias, en la que se encuentran todos los libros, artículos y trabajos que se citan en el texto principal del artículo. Renombrados científicos, al llegar a sus manos el trabajo de un colega, pasan páginas con rapidez convulsa para llegar a la mencionada lista y comprobar si aparecen citados o no. ¿Por qué tanto interés? ¿Pura vanidad? Puede que sea así en buena parte, pero existe otra razón: hoy en día, acertada o desacertadamente, la calidad científica de un trabajo científico se mide por el número de veces que ha sido citado. Aunque cueste reconocerlo, muchos tribunales que juzgan puestos o becas de investigación recurren al número de citas de un artículo para determinar su relevancia, en lugar de leerlo.

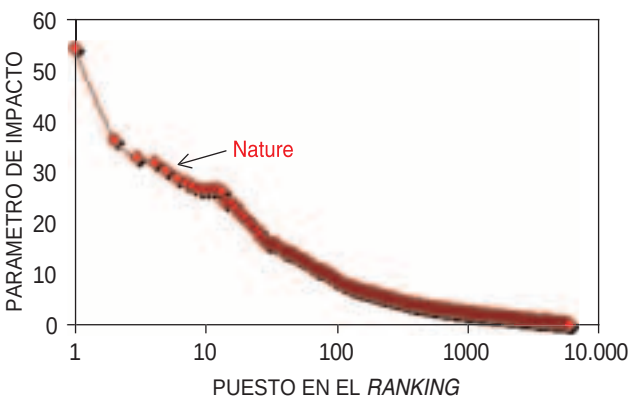
La empresa privada Institute of Scientific Information (ISI) nació en 1958 con el objetivo de recopilar los títulos y resúmenes de todas las publicaciones científicas. En 1961 decidió recopilar también información acerca de las citas. Disponer de un índice en donde se pueden buscar todos los artículos que han citado alguno dado es una herramienta extremadamente útil. Suponga que usted comienza a trabajar en un cierto tema y que encuentra un artículo de 1973 con ideas parecidas a las suyas. Rastrear todos los artículos que, desde 1973 hasta hoy, han citado aquel trabajo le permitirá conocer toda la investigación que se ha hecho en estos últimos 30 años sobre el tema y las ideas que le interesan, algo que es bastante más difícil de obtener sólo mediante búsquedas que hacen uso de las palabras que aparecen en el título o en el resumen de los artículos.

Lo que quizá no esperaban los creadores del ISI es que la recopilación de citas se convertiría con el tiempo en uno de los métodos más utilizados para evaluar, no sólo la calidad de trabajos de investigación, sino también el impacto de las revistas científicas. En efecto, lo que hoy en día determina el éxito de una revista científica es sobre todo su *parámetro de impacto*, que no es más que el número medio de citas que recibe cada artículo publicado en la revista. Una de las revistas con mayor impacto es, por ejemplo, *Nature*. Su parámetro de impacto en 2002 es 30,432 porque los 1952 artículos publicados en los años 2000 y 2001 han sido citados, durante 2002, 59.403 veces, es decir, 30,432 veces por artículo. Este es un número medio de citas bastante alto. De las 5830 revistas registradas en el ISI, *Nature* es la quinta con mayor parámetro de impacto. En la figura 1 se puede ver el parámetro de impacto de estas 5830 revistas, en función del puesto que ocupan en una lista en la que se ordenan de mayor a menor parámetro de impacto.

Si las citas son una buena medida de la importancia de un artículo o una revista es un asunto que está permanentemente en cuestión. En primer lugar, existe una cierta picaresca en torno a las citas, ya que editores y censores de revistas tienden a "recomendar" que se citen sus artículos. En segundo lugar, son muy citados los artículos que contienen alguna idea fácil de generalizar, independientemente de que sea relevante o no, mientras que un artículo que resuelva de una vez por todas un viejo problema puede que se cite muy pocas veces.

A esta polémica se ha sumado una sorprendente investigación realizada por M. V. Simkin y V. P. Roychowdhury, de la Universidad de California en Los Angeles, y que demuestra que los científicos no somos especialmente cuidadosos ni críticos a la hora de citar otros trabajos. En un primer artículo, titulado "¡Leed antes de citar!", investigaron cómo se propagan los errores en las citas: erratas en los nombres de los autores o en la página del artículo citado. Simkin y Roychowdhury analizaron estadísticamente estas erratas en el caso de un artículo citado 4300 veces y comprobaron que los errores se "propagan" a lo largo de las citas. Por ejemplo, de las 45 erratas que detectaron, una de ellas, las más frecuente, aparecía 78 veces, lo cual indica que muchos de los autores que citan ese trabajo han copiado la cita de la lista de referencias de otros artículos sin comprobar los datos del trabajo citado y, lo que es peor, probablemente sin leerlo.

Un argumento muy simple nos puede servir para estimar la fracción de "citadores" que realmente han leído el artículo. Simkin y Roychowdhury detectaron 196 erratas en las 4300 citas. De estas 196, sólo 45 eran diferentes. Es evidente que, de los 196 autores que han citado erróneamente, 196 – 45 = 151 han copiado,



1. Parámetro de impacto de 5830 revistas en función del puesto que ocupan en una lista en donde las revistas se ordenan de mayor a menor parámetro de impacto

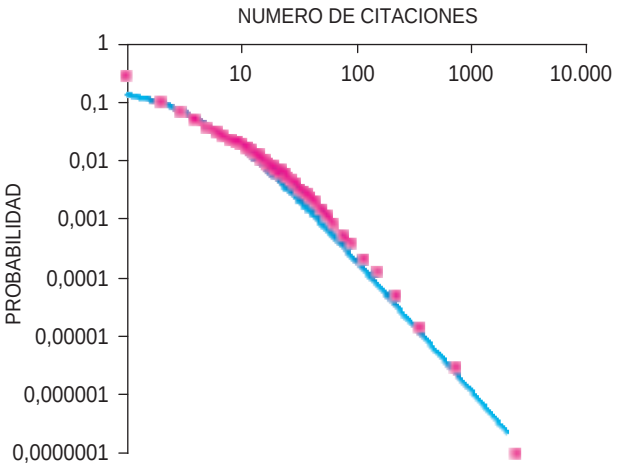
puesto que su errata no es original. Admitiendo la "presunción de inocencia" de los otros 45, es decir, suponiendo que ellos sí han leído el trabajo citado, obtenemos que la fracción de lectores frente a citadores es $R = 45/196 = 0{,}23$. Es decir, según esta deducción, los científicos leen menos de la cuarta parte de los artículos que citan. Un argumento más elaborado basado en un modelo probabilístico conduce a un valor casi idéntico, $R = 0{,}22$.

Si los científicos citan copiando la lista de referencias de otros artículos en lugar de leer las referencias originales, entonces puede que tenga lugar un efecto "bola de nieve" en las citas que recibe un artículo. Simkin y Roychowdhury, en un segundo trabajo, han demostrado que esto es posible. Han diseñado un modelo muy simple que reproduce la estadística de las citas. El modelo podría llamarse de "cita a ciegas", ya que supone que cada científico elige las citas al azar, de acuerdo con las reglas siguientes. Supongamos que se han escrito N artículos y que un científico escribe uno nuevo. El científico escoge al azar m artículos de los N existentes, los cita y cita también, con una probabilidad p, las referencias que aparecen en esos m artículos.

El modelo es tan simple que se puede simular fácilmente con un programa de ordenador, comenzando con m artículos sin lista de referencias. Puede incluso resolverse de modo aproximado, aunque la forma de hacerlo es demasiado complicada para esta sección. En la figura 2 se muestra la solución del modelo con $m = 3$ y $p = 1/4$. Estos son los valores que mejor se ajustan a los datos reales, extraídos del ISI, y que también se muestran en la figura. La gráfica representa la probabilidad de que un artículo haya sido citado k veces, o equivalentemente, la fracción de artículos citados k veces, en función de k. Como se ve en la figura, el modelo reproduce bastante bien los datos reales. Podemos también comprobar el efecto "bola de nieve" mencionado antes. Supongamos que hay N artículos y que uno de ellos ha sido citado k veces. ¿Cuál es la probabilidad de que lo cite el artículo $N + 1$? Recordemos que en el modelo un nuevo artículo elige m al azar entre los N existentes, los cita y cita con probabilidad p los artículos que aparecen en las listas de referencias de los m elegidos. Por tanto, hay dos posibilidades para ser citado: o bien a) el artículo es elegido entre los m primeros, o b) está en la lista de referencias de algunos de estos m primeros artículos. La probabilidad de que ocurra a) es, aproximadamente, m/N y la de que ocurra b) es km/N. Pero, en el caso b) el artículo es sólo citado con una probabilidad p. Por tanto, la probabilidad de que nuestro artículo con k citas sea citado de nuevo es:

$$\frac{m + pkm}{N}$$

y es mayor cuanto mayor sea k. De hecho, si k es pequeño, será muy poco probable que el artículo sea citado de nuevo. En el modelo de cita a ciegas habrá muchos artículos apenas citados y unos pocos muy citados, tal y como muestra la figura 2. Y es ésta la distribución a la que se ajustan los datos reales.



2. La gráfica muestra la probabilidad de que un artículo tenga un número de citas determinado. Sólo aproximadamente uno de cada un millón de artículos es citado más de 1000 veces. La línea azul corresponde al modelo de "cita a ciegas", mientras que los puntos color magenta son los datos reales tomados del ISI. [Figura cedida por M. V. Simkin]

Otra cantidad que puede extraerse fácilmente del modelo es el número medio de citas en cada artículo. Supongamos que, cuando hay N artículos publicados, la lista de referencias de cada uno ellos tiene, de media, $c(N)$ referencias. Un nuevo artículo aparece y, en media, tendrá un número de referencias:

$$c(N + 1) = m + mpc(N)$$

como se deduce fácilmente de las reglas del modelo de cita a ciegas. Cuando el número total de artículos N es muy grande, el tamaño medio de la lista de referencias apenas cambia. Si c es este valor casi constante, entonces tiene que verificar $c = m + mpc$, lo que da un valor para el tamaño medio de la lista de referencias:

$$c = \frac{m}{1 - mp}$$

Esta fórmula es sólo válida si mp es menor que uno. Si no fuera así, las listas de referencias crecerían indefinidamente. Los valores que mejor reproducen los datos reales son, $m = 3$ y $p = 1/4$, lo que da un tamaño medio de $c = 12$ referencias. Los datos del ISI indican que este número es ligeramente superior, en torno a 15. Sin embargo, el acuerdo sigue siendo bueno, especialmente si se tiene en cuenta la simplicidad del modelo.

Aunque es cierto que los científicos muchas veces citan artículos que no han leído, no lo hacen realmente a ciegas, por supuesto. Sin embargo, el modelo sí pone de manifiesto que la red de citas, que puede considerarse como la red en la que se crea y por la que se transmite el conocimiento científico, tiene una estructura bastante homogénea y sorprendentemente sencilla.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La frecuencia fantasma

La mayoría de las lectoras y lectores de *Investigación y Ciencia* probablemente conozcan algunas ilusiones ópticas. Lo que quizá no resulte tan conocido es que también existe un buen número de ilusiones auditivas.

La escala de Shepard y la de Risset, por ejemplo, son los análogos acústicos a las escaleras sin fin de Escher, en las que se puede volver al punto de partida bajando siempre escalones. La escala de Shepard consiste en una serie de notas que, al tocarse de forma cíclica, dan la sensación de formar una escala siempre ascendente o siempre descendente. La escala de Risset es la versión continua de la de Shepard, es decir, un sonido cuya altura parece aumentar indefinidamente, a pesar de ser estrictamente periódico. En Internet se pueden encontrar varias demostraciones de estas escalas y de otras ilusiones acústicas.

Una ilusión acústica más simple es la llamada "fundamental inexistente". Para entender en qué consiste, conviene recordar qué características definen un sonido, especialmente un sonido musical, y cómo se las percibe. Cuando oímos una nota musical, distinguimos entre el timbre y la altura de la nota, es decir, si se trata de un *do* de la escala central del piano o de un *fa* dos octavas más arriba, por ejemplo. Cualquier sonido se puede descomponer en una suma de los llamados *tonos puros*, unas oscilaciones perfectamente periódicas y cuya expresión matemática es del tipo $\mathrm{sen}\,(2\pi\upsilon t)$, en donde υ es la frecuencia del tono puro. La mayoría de los instrumentos musicales emiten sonidos que constan de varios tonos puros llamados armónicos y cuyas frecuencias son múltiplos de una dada que se llama *fundamental*. Por ejemplo, el sonido de una flauta tocando el *la* de la octava central del piano consta de una fundamental de 440 hertz, es decir, una tono puro que oscila 440 veces por segundo, y de sus sucesivos armónicos: $440 \times 2 = 880$, $440 \times 3 = 1320$, $440 \times 4 = 1760$, etc. La intensidad de cada uno de estos armónicos es menor cuanto más alto es el armónico. El timbre de un instrumento está precisamente determinado por las intensidades de estos armónicos. El *la* de una flauta y el *la* de un clarinete constan de la misma fundamental y de los mismos armónicos. Por eso identificamos la misma nota en ambos sonidos. Sin embargo, las intensidades de los armónicos son distintas en cada instrumento y es eso lo que hace que tengan timbres diferentes.

Un sonido como este:

$$\mathrm{sen}\,(2\pi \times 440 \times t) + 0{,}5\,\mathrm{sen}\,(2\pi \times 880 \times t) + 0{,}5\,\mathrm{sen}\,(2\pi \times 1320 \times t)$$

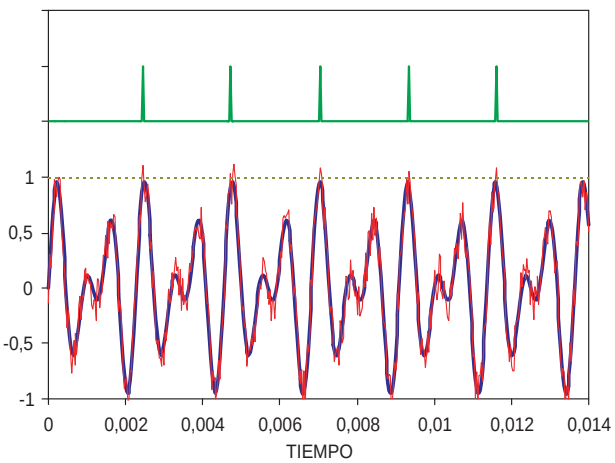
suena parecido a un órgano de iglesia tocando el *la*. Consta de la fundamental de 440 hertz más los dos siguientes armónicos con una intensidad mitad de la intensidad de la fundamental.

Lo sorprendente es que, si eliminamos la fundamental y nos quedamos con:

$$0{,}5\,\mathrm{sen}\,(2\pi \times 880 \times t) + 0{,}5\,\mathrm{sen}\,(2\pi \times 1320 \times t)$$

el oído humano sigue percibiendo este sonido como un *la* de la octava central. Esta es la ilusión de la fundamental inexistente. A pesar de no estar presente en el sonido, la fundamental se puede oír. Esta ilusión acústica es la que nos permite identificar la melodía de una canción en el teléfono. En efecto, la línea telefónica y el micrófono o el altavoz del auricular no son capaces de transmitir o reproducir frecuencias bajas. En consecuencia, lo que oímos al teléfono es un sonido en el que se han eliminado esas frecuencias bajas, entre las que suelen estar las fundamentales de la melodía. Sin embargo, la ilusión de la fundamental inexistente nos permite reconstruir las fundamentales e identificar sin problemas la melodía. Lo sabían hace más de 200 años los constructores de órganos, que en lugar de construir grandes tubos para las notas más bajas, los sustituían por pares de tubos más pequeños cuyo sonido corresponde a los dos primeros armónicos de la nota que se quiere reproducir.

Hay varias teorías que han tratado de explicar esta ilusión acústica. Hace poco más de un mes se ha publicado la más reciente, propuesta por Dante Chialvo. Utiliza una variante de la *resonancia estocástica*, un fenómeno al que en esta misma sección dedicamos el artículo *Ruidos reveladores*, en junio de 2002, y que muestra cómo añadiendo ruido a una señal se puede mejorar la percepción de la misma. Dante Chialvo ha demostrado que la resonancia estocástica puede hacer que se detecte la fundamental inexistente.



1. La curva azul es la representación de un sonido compuesto por dos tonos puros de 880 y 1320 hertz, es decir, el primer y segundo armónicos del la central, de 440 hertz. La curva roja representa este sonido al que se la ha añadido un ruido, es decir, un número aleatorio entre –0,2 y 0,2. En verde se muestra la respuesta de una neurona cuyo umbral de disparo es 1; umbral también representado en la figura por la línea horizontal punteada

En la figura 1 se muestra en azul la representación gráfica del sonido correspondiente a la última fórmula, es decir, el sonido con los dos armónicos de frecuencias 880 y 1320 hertz y sin la fundamental de 440. A este sonido se le añade un ruido, que no es más que un número aleatorio ente –0,2 y 0,2. El resultado es la curva roja de la figura, bastante irregular. La teoría de Chialvo supone que esta señal con ruido llega a una neurona y que ésta responde de la siguiente forma: si la señal con ruido supera cierto umbral (*1 en la figura*), se excita (o dispara). La neurona tarda un cierto tiempo (0,3 milisegundos) en recuperarse y estar en condiciones de volver a excitarse. La curva verde muestra los disparos obtenidos en una simulación de este modelo. Como puede verse en la figura, la neurona, gracias al ruido, se excita en los picos de la onda (aunque no en todos), que están separados exactamente por 1/440 segundos. Es decir, la frecuencia de estos picos y, por tanto, la frecuencia de los disparos de la neurona, es precisamente la de la fundamental inexistente, es decir, 440 hertz. Como vemos, la neurona detecta, gracias al ruido, la frecuencia de la fundamental inexistente.

Sin embargo, el lector suspicaz habrá pensado que, en este ejemplo, el ruido es innecesario. Bastaría reducir un poco el umbral de la neurona, hasta 0,9 por ejemplo, para que detectara los picos a la frecuencia de la fundamental. Para este ejemplo concreto, esa objeción es cierta y, de hecho, la teoría de que la altura del sonido se percibe detectando la frecuencia de los picos de la señal sin necesidad de ruido se remonta a 1940 y se debe a J. F. Schouten. Sin embargo, la propuesta de Chialvo explica peculiaridades de la percepción de la altura de un sonido que sí necesitan el ruido que se añade a la señal.

Cuando se reproduce un sonido que es suma de dos tonos puros que no son armónicos de una fundamental, el oído sigue reconociendo una única nota. En este caso la percepción es más compleja. Hay, por ejemplo, sonidos antes los cuales el oído puede percibir dos notas diferentes. Se percibe una u otra dependiendo de qué es lo que se haya oído antes. La teoría de Chialvo sí puede explicar este fenómeno gracias al ruido. En la figura 2, se comparan los experimentos realizados por Schouten en 1940 con el modelo que acabamos de describir. En el experimento de Schouten, tres individuos escuchan varios sonidos, todos ellos compuestos por tres tonos puros de frecuencias $f - 200$, f y $f + 200$ hertz (estos tres tonos sólo son armónicos de una fundamental si f es múltiplo de 200 hertz) y se les pide que identifiquen a qué nota musical corresponde cada sonido. El resultado se puede ver en la gráfica de la figura 2, en donde se muestran con círculos blancos y azules y triángulos blancos las respuestas de las tres personas. Observen que hay sonidos para los cuales un mismo individuo ha identificado dos notas distintas. Normalmente, si se parte de un sonido compuesto por armónicos de una fundamental, por ejemplo, si se suman tres tonos puros con frecuencias 1400, 1600 y 1800 hertz ($f = 1600$ en la gráfica), el oyente identificará la fundamental inexistente, es decir, 200 hertz,

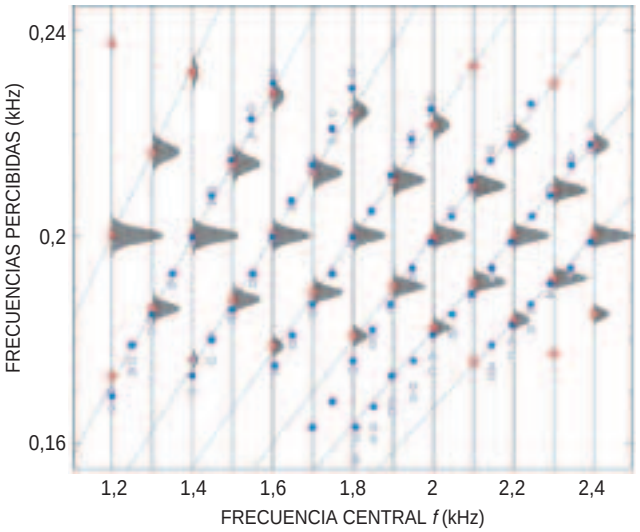


FIGURA CEDIDA POR D. CHIALVO

2. En el experimento de Schouten tres individuos escuchaban un sonido compuesto por tres tonos puros de frecuencias f – 200, f y f + 200 hertz. Los círculos azules y los triángulos y círculos blancos representan las respuestas de los tres individuos ante los sonidos mostrados. Las campanas grises indican la probabilidad con que la neurona de Chialvo se excita a una frecuencia dada y los puntos rojos son los máximos de esta probabilidad.

tal y como muestra la gráfica. Si, a partir de este sonido, aumentamos en paralelo las tres frecuencias, entonces la nota percibida subirá y, si descendemos las frecuencias, descenderá la nota. Pero lo mismo ocurre si partimos de un sonido con $f = 1400$ hertz. Por eso se aprecian en la figura distintas ramas, cada una de las cuales corresponde a una serie de sonidos que se encuentran alrededor de uno compuesto por tres armónicos de 200 hertz y que es percibido como una nota de 200 hertz. Finalmente, las campanas grises indican la probabilidad con la que la neurona propuesta en la teoría de Chialvo se excita a una frecuencia dada y los puntos rojos son los máximos de esta probabilidad. Por lo tanto, los puntos rojos son en realidad las frecuencias que la neurona detecta o puede detectar. No olvidemos que el modelo contiene ruido y, en consecuencia, un cierto grado de aleatoriedad. Como vemos, el acuerdo entre la teoría y el experimento es bastante bueno. En realidad, las distintas ramas del experimento de Schouten pueden reproducirse con otras teorías. Una teoría previa de J. H. Cartwright, D. L. González, y O. Piro, físicos de la Universidad de las Islas Baleares, da lugar a las mismas predicciones que la neurona de Chialvo utilizando osciladores no lineales. La ventaja del modelo de Chialvo es su sencillez y también que permite explicar la ambigüedad en la percepción de la altura del sonido ya que, al introducir aleatoriedad en la percepción, el oyente puede detectar una u otra nota con unas ciertas probabilidades. Tanto este modelo como los que comentamos en el artículo de la resonancia estocástica, indican que el ruido y el azar son cada vez más relevantes para entender la percepción y otros fenómenos cognitivos.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Las ventajas de la solidaridad

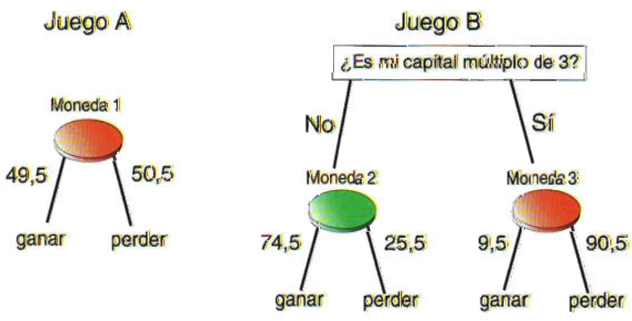
Hace más de dos años que comenzamos esta nueva etapa de la sección de *Juegos matemáticos*, y lo hicimos comentando dos juegos de azar cuyo comportamiento es sorprendente (Juegos matemáticos, julio 2001). En ambos, un individuo juega contra un casino y tiene ciertas probabilidades de ganar y perder un euro en cada turno.

El primero de ellos —lo llamaremos juego A— es similar a apostar un 1 euro a rojo o negro en la ruleta de un casino: ganamos 1 euro con una probabilidad ligeramente inferior al 50 % y perdemos 1 euro con una probabilidad ligeramente superior al 50 %, ya que con el cero gana siempre la banca. Supongamos que se gana con una probabilidad del 49,5 % y se pierde con una probabilidad del 50,5 % (estas probabilidades no coinciden exactamente con las de la ruleta, pero nos sirven para describir la paradoja). Podemos también imaginar este juego como una apuesta sobre el resultado del lanzamiento de una moneda ligeramente sesgada. El segundo juego —lo llamaremos juego B— es un poco más complicado y, obviamente, no puede encontrarse en ningún casino. Igual que en el juego A, en cada turno ganamos o perdemos 1 €, pero ahora las probabilidades dependen de lo que llevamos ganado hasta el momento (que puede ser una cantidad negativa): si lo que llevamos ganado —lo llamaremos *el capital*— es múltiplo de tres, entonces ganamos 1 € con probabilidad 9,5 %; si el capital no es múltiplo de tres, la probabilidad de ganar es del 74,5 %. Es decir, en el juego B se utilizan dos monedas, una muy desfavorable para el jugador, que se lanza cuando el capital es múltiplo de tres, y la otra bastante favorable, que se lanza cuando el capital no es múltiplo de tres. En la figura 1 se esquematizan las reglas de los dos juegos, representándose en rojo las monedas desfavorables y en verde la favorable.

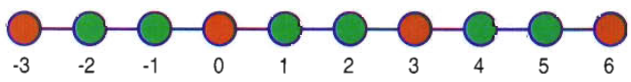
Aunque se trate de juegos de azar, el jugador pierde en media si juega muchos turnos seguidos a cualquiera de los dos juegos, A o B. Sin embargo, si juega alternándolos, ya sea al azar o siguiendo alguna secuencia periódica como AABBAABB..., entonces gana en media. Este comportamiento es a primera vista sorprendente y por ello se conoce a estos juegos como juegos paradójicos o también como *Paradoja de Parrondo*.

En los últimos años, algunos físicos y matemáticos han estudiado variantes de la paradoja original y encontrado nuevas propiedades sorprendentes de estos juegos. En este artículo vamos a comentar una de ellas, que tiene lugar cuando varios individuos juegan a B y se les permite repartir las ganancias entre turno y turno. Raúl Toral, de la Universidad de las Islas Baleares, ha demostrado que un simple reparto del capital puede convertir un juego perdedor en ganador.

Para entender cómo es esto posible, analizaremos primero uno de los mecanismos que explican la paradoja original. Piense por un momento cuántas veces se utiliza la moneda mala cuando jugamos B un gran número de turnos seguidos. A primera vista, parece que la moneda mala se debería utilizar un tercio de los turnos, puesto que se lanza siempre que el capital es múltiplo de tres. Sin embargo, si se juega B en todos los turnos, resulta que la moneda mala se utiliza 5/13 = 0,3846... de las veces que se juega, es decir, más a menudo que un tercio de las veces. La razón se encuentra en las propias reglas del juego B y se hace más clara si representamos el juego como el movimiento de una ficha a lo largo de la línea de la figura 2. Cada vez que ganamos, movemos la ficha una casilla hacia la derecha y, cuando perdemos, una casilla hacia la izquierda. Con las reglas del juego B, cuando la ficha está en una casilla roja, su movimiento más probable es hacia la izquierda, puesto que utilizamos la moneda mala, que tiene una probabilidad de ganar muy reducida, del 9,5 %. Por el contrario, en las casillas verdes el movimiento más probable es hacia la derecha. Si la ficha se encuentra en la casilla 2, por ejemplo, su movimiento más probable es hacia la casilla 3. Pero, en el siguiente turno, lo más probable es que vuelva a la 2. Por tanto, la ficha se encontrará la mayor parte del tiempo saltando entre 2 y 3, o, en general, entre un múltiplo de tres y su inmediato inferior. Esto hace que la frecuencia con la que la ficha visita las casillas rojas, es decir, la frecuencia con la que se



1. Las reglas de los juegos paradójicos A y B. Las monedas que se utilizan en cada juego son rojas o verdes según sean desfavorables o beneficiosas para el jugador



2. Representación esquemática del capital en el juego B

utiliza la moneda mala del juego B, sea superior a 1/3, ya que esta frecuencia es la que correspondería a una ficha que se mueve completamente al azar, sin ninguna preferencia de salto. Pero precisamente este movimiento por completo al azar es el que tiene el capital cuando se juega al juego A. Esta es una de las claves de la paradoja: cuando el juego B se alterna con el A, este último, a pesar de ser perdedor, hace que se utilice menos frecuentemente la moneda mala del juego B y por ello su efecto final es positivo y la alternancia resulta ser ganadora.

El juego A hace que el capital sea más aleatorio de lo que es cuando se juega sólo B; de este modo, la ficha visita menos veces las casillas rojas. Podemos decir que el juego A ayuda a "saltar" estas casillas rojas, en donde el juego B es muy desfavorable.

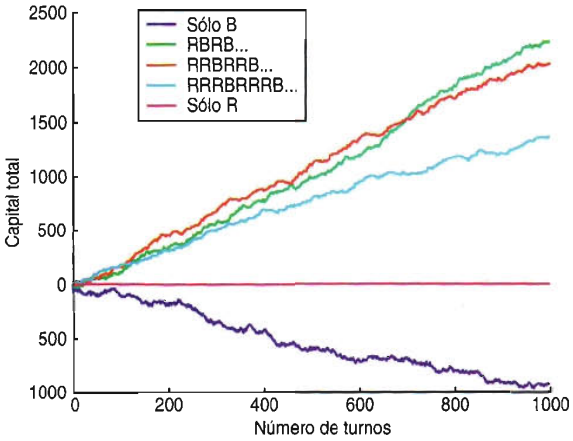
Si varios jugadores están jugando contra la banca, este mismo efecto se puede conseguir de una forma más sencilla: a los jugadores les es posible convertir el juego B en ganador sin más que redistribuir entre turno y turno sus ganancias.

Supongamos N individuos que juegan contra la banca al juego B. Ahora no disponemos del juego A, pero podemos redistribuir el capital entre los distintos jugadores. Llamaremos a esta distribución o reparto juego R. Este juego R cumple el mismo papel que el juego A en la paradoja original. He elegido como forma de redistribución la siguiente: el jugador i le da un euro al $i+1$ con una probabilidad 1/2, y recibe un euro de él con una probabilidad 1/2. Un turno de juego R consiste en realizar estos intercambios para todos los jugadores i desde el 1 hasta el $N-1$ (el jugador N sólo intercambia capital con el $N-1$).

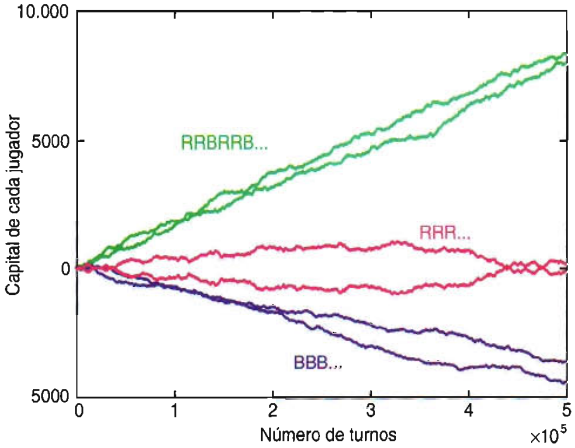
Los resultados de esos juegos se muestran en la figura 3. En esta figura se representa el capital total para 100 jugadores en función del número de turnos jugados y para distintas redistribuciones. La curva azul muestra el capital cuando no se realiza ninguna redistribución, es decir, cuando cada individuo juega contra la banca al juego B sin intercambiar nada con sus compañeros. En ese caso, todos los jugadores pierden en media y eso se ve reflejado en la disminución paulatina del capital total. Sin embargo, si se intercalan turnos de redistribución, como ocurre en las curvas roja, verde y añil, todos los jugadores ganan y el capital total aumenta. Cuando se utilizan muchos turnos en la redistribución, la ganancia se hace menor, aunque sigue siendo positiva. Como en el juego R el capital total ni aumenta ni disminuye (el individual sí), si todos los turnos se utilizan en redistribuir el capital y no se juega nunca al juego B, el capital no variará y el resultado será la recta horizontal morada de la figura. Como vemos, en estos juegos compartir es sin duda beneficioso para todos. Volviendo al esquema de la figura 2, lo que está ocurriendo es que, cuando se comparte el capital, la ayuda del vecino puede facilitarle a un individuo saltar la casilla roja y que se beneficie de las monedas buenas del juego B. Se podría decir que esto es equivalente a superar una "mala racha" con el apoyo de algún amigo o de la colectividad.

El efecto se puede conseguir incluso con sólo dos jugadores. En este caso, la redistribución es muy simple: el primero le da al segundo 1 € con una probabilidad 1/2, o es el segundo quien le da al primero 1 € con probabilidad 1/2 en cada turno de redistribución (juego R). Esta simple redistribución de capital tiene unos efectos cruciales en el desarrollo del juego. En la figura 4 mostramos el capital de cada uno de los jugadores cuando no hay reparto, cuando solamente hay reparto y no se juega nunca a B, y cuando se intercala un turno de juego B con dos turnos de reparto. La redistribución de capital beneficia a ambos, mientras que el juego en solitario es de nuevo perdedor.

El modelo de Toral enseña algo que es cada vez más necesario recordar en los tiempos que corren, especialmente a los defensores del neoliberalismo: la redistribución de la riqueza es beneficiosa para la colectividad e incluso puede ser indispensable para generar crecimiento económico.



3. Capital total de 100 jugadores en función del número de turnos para distintas combinaciones del juego B y la redistribución R



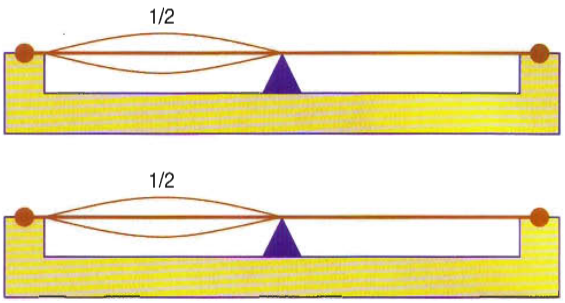
4. Capital de dos jugadores en función del número de turnos cuando juegan sólo a B (curvas azules), *a R* (curvas moradas) *y cuando alternan el juego B y la redistribución R siguiendo la secuencia RRBRRB...* (curvas verdes)

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La teoría matemática de la consonancia

Las matemáticas y la música han estado unidas al menos desde la civilización griega. Los pitagóricos se dieron cuenta de que dos notas tocadas simultáneamente eran agradables si entre ellas existía una relación matemática simple. Al pulsar una cuerda tensada suena una nota, que depende de la tensión, la longitud y la densidad de la cuerda. Si disminuimos la longitud de la cuerda a la mitad, sin variar su tensión, por ejemplo, colocando una cuña como en la figura 1, sonará una nota más alta que la primera, pero la combinación de ambas será agradable al oído. Es más, la segunda nota suena en realidad muy parecida a la primera. El intervalo entre estas dos notas se denomina *octava* y su sonido es tan parecido, que en música las dos notas se denominan con el mismo nombre. Si la primera es Do, por ejemplo, la segunda nota será otro Do, aunque en una octava superior.



1. Los primeros intervalos pitagóricos. Al dividir la longitud de una cuerda por la mitad se pasa de una nota Do, por ejemplo, al Do de la octava superior. Al disminuir la cuerda hasta 2/3 de su longitud original se pasa del Do al Sol.

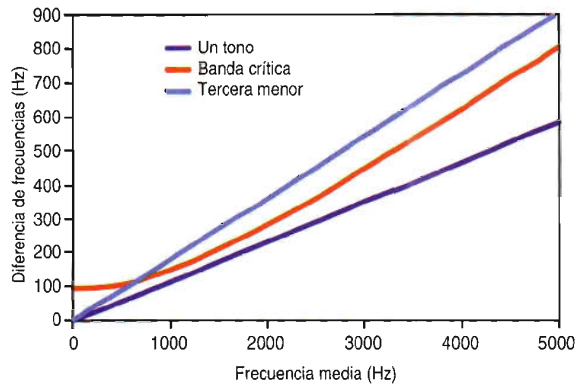
Pero no sólo el intervalo de octava es agradable musicalmente. Los griegos descubrieron que, si la cuerda se acorta hasta 2/3 de su longitud original, la nueva nota también se combina agradablemente con la nota correspondiente a la cuerda completa (*véase la figura 1*). Suponiendo de nuevo que la nota original es Do, la nota de la cuerda 2/3 más corta sería Sol. Hay algunos otros intervalos agradables, que técnicamente se llaman *consonantes*, mientras que otros, la mayoría, son desagradables o *disonantes*.

No existe una clasificación nítida entre intervalos consonantes y disonantes. Sí se da un grado de consonancia o disonancia, que incluso depende del instrumento en el que se toca el intervalo o de la altura de las notas que lo forman. Por ejemplo, un intervalo Do-Mi es claramente consonante; ahora bien, si estas dos notas se tocan en las escalas más bajas de un piano, el sonido que resulta no es del todo agradable. La consonancia también depende, salvo en casos muy claros, como la octava, de la cultura y la educación musical del oyente.

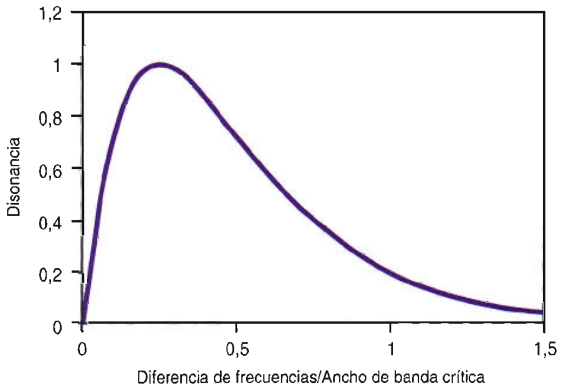
Tanto la consonancia como la disonancia de intervalos constituyen la base de la armonía en la música occidental. Gracias a ellas se construyen las tensiones y las resoluciones que forman el discurso musical. Sin embargo, no hemos tenido una teoría satisfactoria de la consonancia hasta hace poco. Fueron Plomp y Levelt quienes, en 1965, elaboraron la teoría más aceptada, después de revisar ideas de Galileo, Leibniz, Euler y Helmholtz, entre otros.

La nota de un piano, de una flauta o una guitarra, es la suma de varios *tonos puros*. Un tono puro es una onda sinusoidal: una vibración del aire cuya expresión matemática es $A\,\mathrm{sen}\,(2\pi\nu t)$, en donde A es la *amplitud* del tono puro y ν su frecuencia, es decir, el número de oscilaciones por segundo. La nota La central del piano está compuesta por un tono puro de frecuencia 440 hertz (440 oscilaciones por segundo) y sus sucesivos múltiplos o armónicos, es decir, tonos de frecuencias $2 \times 440 = 880$ hertz, $3 \times 440 = 1320$ hertz, etc. (véase *Juegos Matemáticos* de enero de 2004). El tono puro de frecuencia más baja, o *fundamental*, determina la altura del sonido que percibimos. Por otra parte, en el caso de una cuerda pulsada, la frecuencia fundamental es inversamente proporcional a la longitud de la misma. De modo que el intervalo de octava lo forman dos notas de frecuencias fundamentales ν y 2ν, y el intervalo de quinta consiste en dos notas de frecuencias ν y $3\nu/2$. De ello se deduce también que el intervalo que forman dos notas no depende de la diferencia de sus frecuencias sino de su cociente, un hecho de especial trascendencia en la teoría matemática de la música.

Las teorías previas de la consonancia se basaban, o bien en la vieja idea pitagórica de que el universo, desde los planetas hasta la armonía musical, se rige por relaciones entre números enteros pequeños, o bien en el hecho de que algunos intervalos consonantes están formados por notas que comparten muchos armónicos. La teoría de Plomp y Levelt mejora y completa esta última idea. En primer lugar, investigaron la consonancia de dos tonos puros de distintas frecuencias, ν_1,ν_2, haciendo que varios oyentes los escucharan y evaluaran en una escala de 1 a 7 el desagrado que les producía. Descubrieron que dos tonos puros cuyas frecuencias estén suficientemente alejadas suenan siempre bien. Más precisamente, existe una *banda crítica* de frecuencias, de modo que, si la diferencia entre las frecuencias de los dos tonos puros excede el ancho de esta banda crítica, entonces los dos tonos son consonantes. El ancho de la banda crítica depende a su vez de la frecuencia media, $\bar{\nu} = (\nu_1 + \nu_2)/2$, de los dos tonos puros presentados, tal y como se aprecia en la figura 2. No disponemos de una expresión matemática para la banda crítica, puesto que se obtiene mediante



2. Anchura de la banda crítica en función de la frecuencia media de dos notas $\bar{\nu} = (\nu_1 + \nu_2)/2$ y comparada con un intervalo de un tono (como el de Do a Re) y uno de tercera menor (como el de La a Do).



3. Disonancia de dos tonos puros en función de la diferencia de sus frecuencias dividida por la anchura de la banda crítica.

experimentación con individuos. Sin embargo, una buena aproximación viene dada por la ecuación:

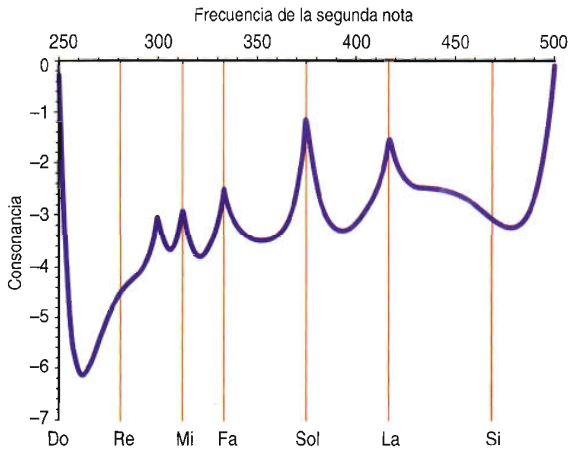
$$\Delta\nu_{crit} = 2\bar{\nu}/11 + 200 \times e^{-\bar{\nu}/1000} - 105$$

Plomp y Levelt también cuantificaron la disonancia dentro de la banda crítica. El resultado se muestra en la figura 3, en donde se observa que la diferencia de frecuencias más disonante es un cuarto de la anchura de la banda crítica. De nuevo la curva es puramente experimental, pero se puede aproximar por la función:

$$d(\nu_1, \nu_2) = 4|x|e^{1-4|x|}$$

en donde $x = (\nu_1 - \nu_2)/\Delta\nu_{crit}$.

Consideremos ahora dos sonidos compuestos por varios armónicos. Tomaremos la fundamental del primer sonido de 250 hertz y variaremos la del segundo entre 250 y 500 hertz, es decir, una octava completa. Supondremos que cada sonido está compuesto por la



4. La curva de consonancia para dos notas. La primera es un Do de 250 Hz y la frecuencia de la segunda varía desde 250 hasta 500 Hz a lo largo del eje horizontal de la gráfica. En rojo se señalan las siete notas de la escala de Do mayor.

fundamental y cinco armónicos, todos ellos de igual amplitud. Para calcular la consonancia o disonancia de estos dos sonidos, basta calcular la disonancia entre los tonos puros que los forman utilizando la fórmula anterior y sumar todas las parejas de tonos puros. El resultado se muestra en la figura 4, en donde también se representan mediante barras rojas verticales las frecuencias de las siete notas de la escala justa de Do mayor. Estas notas se relacionan con el Do inicial mediante los siguientes factores: Do (1), Re (9/8), Mi (5/4), Fa (4/3), Sol (3/2), La (5/3) y Si (15/8). (En realidad, el Do central del piano tiene una frecuencia fundamental de 261,6 Hz, y no de 250 Hz, como se deduce de la gráfica, pero he elegido esta última frecuencia para simplificar la exposición). En lugar de la disonancia entre los dos sonidos, he dibujado la consonancia, que no es más que la disonancia cambiada de signo. Se puede observar cómo aparecen picos en los intervalos consonantes Do-Mi, Do-Fa, Do-Sol, Do-La, y Do-Do (hay también un sexto pico que coincide con el Mi bemol). El lector puede intentar reproducir la curva de consonancia para otros sonidos, formados por otros armónicos con distintas amplitudes. De hecho, incluso se puede encontrar una aplicación Java en Internet que dibuja estas gráficas para diferentes instrumentos, así como explicaciones más detalladas del fenómeno. Especialmente recomendable es el libro *Mathematics and Music*, de Dave Benson, de la Universidad de Georgia, que se puede descargar gratuitamente desde su página web.

De todos modos, la teoría de Plomp y Levelt no explica completamente la consonancia y disonancia de intervalos. John R. Pierce, autor de otro libro delicioso que también habla de la consonancia, *Los Sonidos de la Música* (Biblioteca Scientific American), demostró que la consonancia de ciertos sonidos formados por tonos puros que no son armónicos pero que producen un valor bajo de disonancia, no es tan "musical" como la consonancia de los intervalos consonantes de la escala de Do mayor. La explicación de que el acorde Do-Mi-Sol, por ejemplo, sea tan agradable al oído, se encuentra no sólo en la ausencia de tonos puros disonantes sino también en que las tres notas son armónicos de una misma fundamental: el Do situado dos octavas más abajo.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Cuestión de escala

La música y las matemáticas tienen mucho en común, especialmente en Occidente. Ya comentamos el mes pasado en esta misma sección el tratamiento matemático de la consonancia y la disonancia entre dos notas. Otro de los problemas en donde las matemáticas tienen un papel crucial es en el diseño de escalas musicales. Una escala es una serie de notas con las que se hacen las melodías de las canciones y piezas musicales que escuchamos todos los días. Pero esta serie de notas es el resultado de una evolución muy compleja, que se explica en parte por razones físicas, pero que además está jalonada por decisiones más o menos arbitrarias que tomaron músicos y constructores de instrumentos de distintas épocas.

Recordemos cuáles son los pares de notas o intervalos consonantes, es decir, los más agradables al oído y los que, por tanto, deberían encontrarse en cualquier escala musical. El más consonante de todos es la octava, el intervalo formado por dos notas de frecuencias f y $2f$. De hecho, dos notas con esas frecuencias suenan prácticamente igual y por eso en música tienen el mismo nombre. Por ejemplo, el La central del piano es un sonido que vibra 440 veces por segundo. Decimos que su frecuencia es de 440 hertz. Un sonido que vibre a 220 hertz es también un La, una octava más abajo, y otro que vibre a 880 hertz es un La una octava más arriba. Todas estas notas suenan tan parecidas, que una melodía se puede tocar en cualquier octava o en varias octavas a la vez y ser perfectamente reconocible. Otros intervalos consonantes son la quinta, que consta de dos notas cuyo cociente de frecuencias es 3/2, la cuarta, definida por un cociente igual a 4/3, la tercera mayor de 5/4, etc. Hay dos aspectos relevantes en la definición de intervalo y de intervalo consonante. En primer lugar, el intervalo entre dos notas está determinado por el cociente de sus frecuencias y no por la diferencia. En segundo lugar, los intervalos consonantes suelen estar formados por el cociente de números enteros sencillos (aunque ya vimos el mes pasado que la teoría de la consonancia es bastante más compleja). Ambos aspectos están determinados por nuestro sistema auditivo y por el modo como asignamos notas a los sonidos que escuchamos.

Diseñar una escala no es más que colocar varias notas entre una nota dada, por ejemplo el Do, y su octava superior. En la música occidental, se utiliza la llamada *escala igual temperada* o simplemente escala temperada. Aunque tiene un sencillo fundamento, dividir la octava en doce intervalos iguales, su historia es algo tortuosa. La razón es que la escala temperada es una solución aproximada a un problema matemáticamente irresoluble: cómo diseñar una escala en la que todos los intervalos sean consonantes. Veamos por qué.

Una escala que se construye a partir de intervalos consonantes es la llamada *escala justa*. La escala justa de siete notas es:

Nota	Frecuencia (con respecto al Do)
Do	1
Re	9/8 = 1,125
Mi	5/4 = 1,25
Fa	4/3 = 1,333
Sol	3/2 = 1,5
La	5/3 = 1,667
Si	15/8 = 1,875
Do	2

Se construye de modo que los principales acordes de tres notas estén formados por intervalos consonantes. Se comienza con la nota Do y se añade el Mi, una tercera mayor por encima (5/4) y el Sol, una quinta por encima (3/2). Después se construye el acorde de Sol subiendo al Si mediante una tercera mayor ($3/2 \times 5/4 = 15/8$), y al Re mediante una quinta ($3/2 \times 3/2 = 9/4$). Finalmente, se construye el acorde de Fa de modo que la tercera nota sea el Do inicial. Así, el Fa estaría una quinta por debajo del Do, es decir, su frecuencia sería de 2/3. Para colocar el Fa dentro de la escala, subimos una octava multiplicando por dos y obtenemos la frecuencia de 4/3. A partir de ese Fa subimos una tercera mayor y obtenemos el La ($4/3 \times 5/4 = 5/3$). La quinta por encima del Fa tiene una frecuencia $4/3 \times 3/2 = 2$ que es el Do de la octava superior. La escala justa tiene los tres acordes de Do, Sol y Fa perfectos, formados por intervalos completamente consonantes. Sin embargo, el intervalo entre Re y La no es una quinta, sino que vale 5/3:9/8 = 40/27 = 1,481, un valor ligeramente inferior a los 3/2 de la quinta perfecta. Esto hace que el acorde de Re suene un poco desafinado en la escala justa.

Los intervalos de quinta son tan importantes, que existe una escala construida a partir de ellos, se trata de la *escala pitagórica*. La diseñaron los pitagóricos al descubrir la consonancia y la relación matemática sencilla de la quinta y de la octava. La escala pitagórica de siete notas tiene las frecuencias siguientes:

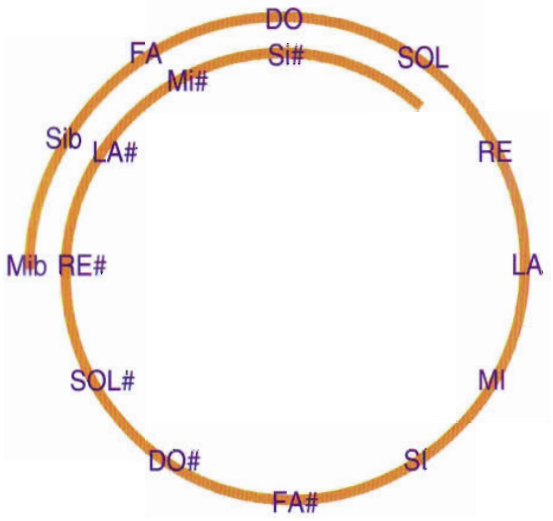
Nota	Frecuencia (con respecto al Do)
Do	1
Re	9/8 = 1,125
Mi	81/64 = 1,266
Fa	4/3 = 1,333
Sol	3/2 = 1,5
La	27/16 = 1,6875
Si	243/128 = 1,898
Do	2

Para construirla es muy útil el llamado círculo de quintas que, en realidad, es una espiral que cerraremos más tarde en forma de círculo. La figura muestra el círculo de quintas. Si lo recorremos en el sentido de las agujas del reloj, cada nota está una quinta por encima de la precedente, es decir, la frecuencia de cada nota es 3/2 de la precedente. La escala pitagórica se construye a partir del círculo "bajando" las notas tantas octavas como sea necesario para que queden dentro de la escala. Por ejemplo, la frecuencia del Re del círculo es $3/2 \times \times 3/2 = 9/4$, que está entre 2 y 4 y queda, por tanto, una octava por encima de nuestra escala. Bajamos dividiendo por dos (recordemos que para subir una octava tenemos que multiplicar la frecuencia por 2 y para bajarla, dividir por 2) y obtenemos los 9/8 del Re de la escala. Las notas de la escala pitagórica tienen frecuencias de la forma $(3/2)^n/2^m$, en donde n y m son números enteros; o bien, si recorremos la espiral en el sentido contrario a las agujas del reloj, $(2/3)^n \times 2^m$.

La escala pitagórica tiene algunos problemas. En primer lugar, ciertos intervalos de la escala no son consonantes. Por ejemplo, entre Do y Mi no hay una tercera mayor (5/4). En segundo lugar, la espiral nunca se cierra porque no existe ningún par de enteros n y m tales que $(3/2)^n = 2^m$, es decir, nunca se vuelve a un Do al recorrer la curva. Sin embargo, en ocasiones las notas de la espiral pasan muy cerca unas de otras. Eso ocurre, por ejemplo, cada 12 notas del círculo. En efecto, $(3/2)^{12} = 129{,}75$ es muy parecido a $2^7 = 128$. Recorrer la espiral de quintas 12 puestos hacia la derecha nos llevaría a un Si sostenido (*véase la figura*), que es prácticamente igual a un Do. El intervalo entre este Si sostenido y el Do más cercano:

$$\left(\frac{3}{2}\right)^{12} \Big/ 2^7 = \frac{3^{12}}{2^{19}} = 1{,}0136$$

es bastante pequeño, más o menos un noveno de tono y se llama *coma pitagórica*.



La espiral pitagórica de quintas es ilimitada en las dos direcciones. A lo largo de la curva, las notas están separadas siempre por una quinta. El mérito de la escala temperada consiste en cerrar la espiral en un círculo.

Los problemas de las escalas justa y pitagórica, y de muchas otras que se diseñaron en los siglos XVII y XVIII, no son excesivamente graves si se toca con ellas una pieza musical que no cambia de tonalidad. En los ejemplos que hemos visto, una pieza en Do sonará razonablemente bien en cualquier escala. Sin embargo, los músicos barrocos querían dar más riqueza a su música pasando de una tonalidad a otra a lo largo de una misma pieza. Para ello se diseñó la escala temperada, que opta por dividir la octava en 12 intervalos iguales o semitonos. Dos notas separadas por un semitono tienen frecuencias cuyo cociente es $2^{1/12}$. Si subimos 12 semitonos a partir de una nota, es decir, si multiplicamos su frecuencia por $2^{1/12}$ doce veces, habremos multiplicado por 2, es decir, habremos subido una octava. La escala temperada de siete notas viene dada por la siguiente tabla:

Nota	Frecuencia (con respecto al Do)
Do	1
Re	$2^{1/6} = 1{,}122$
Mi	$2^{1/3} = 1{,}26$
Fa	$2^{5/12} = 1{,}335$
Sol	$2^{7/12} = 1{,}498$
La	$2^{9/12} = 1{,}682$
Si	$2^{11/12} = 1{,}888$
Do	2

La escala temperada tiene también la virtud de cerrar la espiral de quintas. En ella el Si sostenido es exactamente igual que el Do, como podemos observar en el teclado de un piano. Aun así, conviene recordar que la escala temperada es una solución "de compromiso" a un problema irresoluble.

Esta es sólo una parte minúscula de la historia de las escalas. Una historia en la que han intervenido músicos y científicos como Galileo o Euler. Existe en Holanda un instituto enteramente dedicado a este tema, la Huygens-Fokker Foundation. Una de sus actividades más espectaculares consiste en el desarrollo del programa informático *Scala*, con el que se pueden diseñar escalas a partir de 11 algoritmos, entre los que se incluyen el pitagórico y el temperado. El programa dispone también de más de tres mil escalas diferentes, todas ellas utilizadas en la historia de la música occidental o en la de otras culturas.

Las escalas son también de gran importancia en la música contemporánea, que está constantemente explorando nuevos diseños, más allá de la escala temperada de 12 notas. Finalmente, para el lector interesado, recomiendo de nuevo el excelente texto de Dave Benson, *Mathematics and Music*, que se puede descargar gratuitamente en Internet. Esta obra dedica más de 70 páginas a la teoría de las escalas.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Matemáticas electorales

Los días siguientes a las elecciones generales siempre se oye y se lee acerca de las arbitrariedades y agravios del sistema electoral: cómo un diputado requiere menos votos en determinadas provincias, cómo tal formación tiene cinco veces más votos que otra pero la mitad de diputados, etc. Las críticas giran sobre todo en torno a la *regla D'Hont*. Señala ésta cómo asignar los escaños de cada provincia a los partidos políticos según el número de votos recibidos. La regla D'Hont es una de las posibles soluciones del *problema de reparto*, que encontramos en cualquier situación en la que deba repartirse algo formado por unidades indivisibles (los diputados del congreso).

Supongamos que tres partidos políticos, A, B y C, han obtenido, respectivamente, 300, 140 y 60 votos, en una provincia a la que corresponden cinco diputados. La fracción de votos del partido A es 300/500 = = 0,6. Su representación en el parlamento debería ser proporcional; es decir, lo más justo es que alcanzara $0,6 \times 5 = 3$ diputados. Hasta aquí el reparto va bien. Pero la fracción de votos del partido B es 140/500 = = 0,28 y le corresponden $0,28 \times 5 = 1,4$ diputados. Finalmente, al partido C le corresponden $(60/500) \times 5 =$ = 0,6 diputados.

Ahora bien: no puede haber diputados fraccionarios. Hay que aproximar el reparto a números enteros. (Podría considerarse un parlamento en el que el valor del voto de los diputados fuera fraccionario, pero semejante posibilidad supondría una complicación excesiva en el funcionamiento de la cámara). El número fraccionario de diputados que le correspondería a cada partido se denomina *cuota de reparto*. En nuestro ejemplo, la cuota de reparto de A es 3, la de B es 1,4 y la de C es 0,6. El problema del reparto consiste en diseñar un método que aproxime estos números a números enteros. Podríamos pensar en redondearlos, procedimiento habitual en las ciencias experimentales: si la parte decimal de la cuota de reparto es inferior a 0,5, redondeamos al entero inferior y si es superior a 0,5, redondeamos al entero superior. Con este método, al partido A le corresponderían 3 diputados, al B 1 diputado y al C 1 diputado. Pero ocurre que no siempre la suma de las cuotas redondeadas es igual al número de diputados a repartir.

Veámoslo. Con los resultados de la tabla siguiente, si pretendemos repartir cinco escaños:

PARTIDO	NUMERO DE VOTOS	CUOTA DE REPARTO	ESCAÑOS
A	230	2,3	2
B	140	1,4	1
C	130	1,3	1

hay un escaño que queda sin repartir. El redondeo de las cuotas no resuelve, pues, el problema del reparto.

Un método que sí lo soluciona es el de Hamilton, o de las partes decimales mayores. En este método, se asigna a cada partido un número de escaños igual a la parte entera de su cuota. Tras ese primer reparto, quedarán por asignar algunos escaños; se atribuirán, uno a uno, a los partidos cuya cuota alcance la mayor parte decimal. Con los datos iniciales, el método de Hamilton funcionaría de la siguiente forma:

PARTIDO	NUMERO DE VOTOS	CUOTA DE REPARTO	PARTE ENTERA	ESCAÑOS
A	300	3,0	3	3
B	140	1,4	1	1
C	60	0,6	0	1

En la primera asignación hemos repartido sólo 4 escaños, 3 para A y 1 para B. El escaño que sobra se asigna al partido cuya cuota tenga la mayor parte decimal. En este caso es el C, cuya parte decimal es 0,6, superior a 0,0 (parte decimal de la cuota de A) y a 0,4 (parte decimal de la cuota de B). El método de Hamilton parece justo a primera vista. Sin embargo, presenta una grave deficiencia. Para sacarla a la luz, veamos qué ocurre si, en vez 5 escaños, se trata de repartir 12. La tabla es entonces (recordemos que la cuota de un partido con x votos es ahora $(x/500) \times 12$):

PARTIDO	NUMERO DE VOTOS	CUOTA DE REPARTO	PARTE ENTERA	ESCAÑOS
A	300	7,2	7	7
B	140	3,36	3	3
C	60	1,44	1	2

El escaño que queda sin repartir con las partes enteras se va al partido C, porque tiene la cuota con parte decimal mayor (0,44). Si repartimos 13 escaños, la tabla sería:

PARTIDO	NUMERO DE VOTOS	CUOTA DE REPARTO	PARTE ENTERA	ESCAÑOS
A	300	7,8	7	8
B	140	3,64	3	4
C	60	1,56	1	1

ya que los escaños sobrantes pasan a los partidos A y B. Para nuestra sorpresa, el partido C posee ahora menos escaños que antes, pese a haber aumentado el nú-

mero total de escaños. Tal situación, conocida por *paradoja de Alabama*, se dio en un estudio sobre la asignación de escaños a los distintos estados de EE.UU. en 1880, lo que condujo al congreso a adoptar otro método de reparto en 1901. Sin embargo en España el reparto de escaños por provincias sigue haciéndose mediante el método de Hamilton. Finalmente, el método de Hamilton adolece de un segundo inconveniente. En ocasiones, el aumento de votos de un partido, manteniéndose constante el número de votos del resto, puede determinar que dicha formación política vea mermada su representación parlamentaria: se trata de la *paradoja de la población*. Lo deseable sería un método exento de ambas paradojas.

La famosa regla D'Hont, que en EE.UU. se conoce como método de Jefferson, se halla libre de tales deficiencias. La regla D'Hont consiste en elegir un común divisor d que divide el número de votos conseguido por cada partido. Los números resultantes se redondean por abajo, es decir, eliminando la parte decimal. Se obtiene, así, un número de escaños asignados cuyo total puede o no ser igual al número total de escaños a repartir. Se ajusta entonces el divisor d, de modo que coincidan ambos números. ¿Cómo actúa en nuestro ejemplo si queremos repartir cinco escaños? Probemos, por ejemplo, los tres divisores $d = 50$, 75, 100 y calculemos los cocientes:

PARTIDO	NUMERO DE VOTOS	$d = 50$	$d = 75$	$d = 100$
A	300	6	4	3
B	140	2,8	1,87	1,4
C	60	1,2	0,8	0,6

Vemos que el divisor 50 da lugar al reparto 6, 2, 1, es decir, un total de 9 escaños, que es mayor que los cinco escaños a repartir. El divisor 100, por el contrario, resulta demasiado alto, ya que conduce al reparto 3, 1, 0, cuyo total es 4. El divisor 75 produce el reparto deseado: 4,1,0, que suman un total de cinco escaños.

Pero el método D'Hont tiene un serio problema. Recordemos que la cuota de reparto del partido A es exactamente 3. Sin embargo, la regla D'Hont le ha otorgado 4 escaños. Lo deseable sería que cualquier método asignase a todos los partidos un número de escaños que fuese o bien el entero inferior a su cuota o el entero inmediatamente superior a su cuota. Por ejemplo, si la cuota es 2,9, el partido debería obtener como máximo 3 escaños y como mínimo 2. Esta propiedad se llama *condición de la cuota* y parece lógico exigir que cualquier método justo la cumpla.

Ahora bien, el método D'Hont no la satisface, según hemos visto. ¿Deberíamos por ello desecharlo? La respuesta no es sencilla. Balinski y Young demostraron en 1960 que no existe ningún método que verifique la condición de la cuota y que esté exento de las paradojas de la población y de Alabama. Estos matemáticos demostraron, por tanto, que no existe un método de reparto perfecto. Hay que sacrificar siempre algo y ese sacrificio constituye ya una decisión política. Muchos piensan que las paradojas de Alabama y de la población son más dañinas que la condición de la cuota y, en consecuencia, defienden el método D'Hont y otros métodos que utilizan un divisor común.

Los métodos del divisor común se basan en el mismo principio que el D'Hont. Se dividen los votos por un divisor común d. Los métodos difieren en la forma de redondear los cocientes obtenidos. El *método de Webster*, por ejemplo, los redondea del modo habitual, tomando el entero superior si la parte decimal del cociente es mayor o igual que 0,5. En nuestro ejemplo, el método de Webster daría un reparto 6,3,1 para $d = 50$; 4,2,2 para $d = 75$; y 3,1,1 para $d = 100$. El divisor correcto sería 100 y el reparto final, 3,1,1, diferente del reparto D'Hont. En este caso, el método de Webster sí cumple la condición de la cuota, pero se pueden encontrar otros ejemplos en donde no la verifica.

De los métodos del divisor, ¿cuál es mejor? Como ya hemos dicho, ninguno es perfecto. Sin embargo, puede demostrarse que el D'Hont favorece los partidos con mayor número de votos (en nuestro ejemplo este sesgo es evidente). De hecho, en EE.UU. lo propuso Jefferson porque su estado, Virginia, era el más poblado en 1790. Este sesgo hacia partidos grandes ha sido muy controvertido en nuestro país. Por otro lado, también Balinski y Young demostraron en 1980 que el único método que carece de sesgos hacia partidos con más, o con menos, votos es precisamente el de Webster. Quizás éste último sea un método de reparto a considerar para nuestro sistema electoral.

Algunos lectores quizá se hayan sorprendido de la exposición que hemos hecho aquí de la regla D'Hont. Habitualmente se explica del siguiente modo. Se divide el número de votos de cada partido por 1, 2, 3, ..., n, siendo n el número de escaños a repartir. De todos esos cocientes se eligen los n mayores y se obtiene así el reparto (en nuestro país, además, se eliminan antes de hacer esta tabla los partidos con un porcentaje de votos inferior al 3 %). En nuestro ejemplo la tabla de cocientes sería:

PARTIDO	NUMERO DE VOTOS	votos/2	votos/3	votos/4	votos/5
A	300	150	100	75	60
B	140	70	46,67	35	28
C	60	30	20	15	12

Los cinco cocientes mayores son: 300, 150, 140, 100 y 75. Cuatro de ellos en la fila del partido A y 1, en la del B. Por tanto, el reparto es 4,1,0, que coincide con el que hemos hallado utilizando el método del divisor. El lector puede demostrar que esta forma de hallar el reparto D'Hont es equivalente a la del divisor.

Hemos visto sólo una pequeña parte de las matemáticas electorales. Si desean profundizar más en ellas, les recomiendo la página web "La elección social: un sueño imposible", de Bartolomé Barceló, profesor de la Universidad Autónoma de Madrid (http://www.uam.es/bartolome.barcelo/curso.html).

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La paradoja del autostopista

Supongamos que se lanza un dado no trucado un gran número de veces y que apostamos un euro a que saldrá un cinco. El pago justo de esta apuesta sería de 6 euros, puesto que la probabilidad de que el resultado de una tirada sea cinco es 1/6. Por pago entendemos lo que el jugador recibe a cambio del euro que depositó, es decir, recibe 6 euros en total o, si lo prefieren, recibe 5 más el euro que depositó.

Un posible resultado del juego es el que se ve en la figura 1. He escrito en rojo todos los cincos que aparecen en la secuencia de tiradas. Una cantidad interesante desde el punto de vista probabilístico es el número de tiradas que separan un cinco del siguiente. Todo aficionado al parchís, y a cualquier juego de azar, sabe que esta cantidad es sumamente caprichosa. En ocasiones podemos esperar muchos turnos hasta que aparezca el número deseado, mientras que otras veces puede aparecer en tres o hasta en cuatro tiradas seguidas. En la secuencia de la figura 1, por ejemplo, entre el primer y segundo cinco hay 8 turnos, entre el segundo y el tercero, 7 turnos, y así sucesivamente. Las respectivas distancias se indican en azul, encima de la llave que une dos cincos consecutivos.

¿Cuál es la probabilidad de que esa distancia sea, digamos, igual a 4? Para que entre un cinco y el siguiente haya 4 turnos, es necesario que no aparezca un cinco durante tres turnos, y que aparezca uno precisamente en el cuarto turno. La probabilidad de que esto ocurra es el producto de las probabilidades de cada uno de los eventos, es decir:

$$\text{Prob [distancia = 4]} = \frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}$$

ya que la probabilidad de que no salga el cinco es 5/6 y la de que sí salga es 1/6. En general, la probabilidad de que dos cincos consecutivos estén separados por n turnos es:

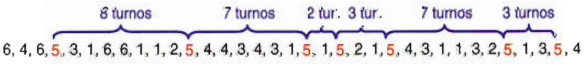
$$\text{Prob [distancia = } n\text{]} = \left(\frac{5}{6}\right)^{n-1}\times\frac{1}{6}$$

En la figura 2 se puede ver una gráfica de esta probabilidad en función de la distancia n (puntos azules). En contra de lo que a primera vista cabría esperar, la distancia más probable es de sólo un turno. Sin embargo, el valor medio de esa distancia, n, es igual a 6. Este valor medio puede calcularse sumando todas las posibles distancias pesadas con su correspondiente probabilidad:

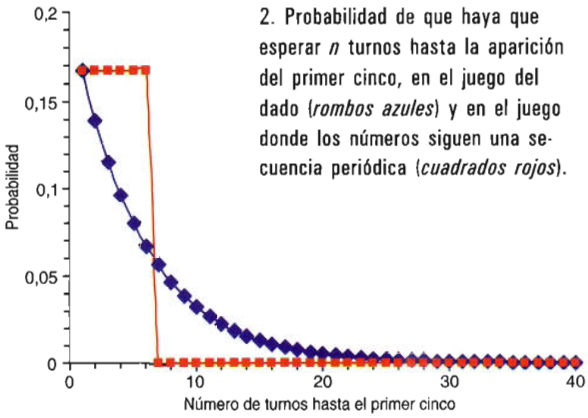
$$\langle n\rangle = 1\times\frac{1}{6} + 2\times\frac{5}{6}\times\frac{1}{6} + 3\times\left(\frac{5}{6}\right)^2\times\frac{1}{6} + \cdots$$
$$\cdots + n\times\left(\frac{5}{6}\right)^{n-1}\times\frac{1}{6} + \cdots$$

pero resulta una suma con infinitos términos que es difícil de evaluar. Una forma sencilla de calcular el valor medio $\langle n\rangle$ es la siguiente. Si jugamos un gran número de turnos, pongamos 6000, el cinco aparecerá, aproximadamente, un sexto de las veces, es decir, unas 1000 veces. La suma de las distancias entre cincos consecutivos es igual a la distancia entre el primer y el último cinco. Esta suma es sólo ligeramente inferior a 6000, de modo que la distancia media es, aproximadamente 6000/1000 = 6. El error cometido con esta aproximación se hace cada vez más pequeño si aumentamos el número de turnos jugados, de modo que, para un número infinito de turnos, el valor medio será exactamente 6.

Supongamos ahora que usted llega a la mesa de juego con la partida ya empezada. ¿Cuál es entonces el número medio de tiradas que tiene que esperar hasta que salga el primer cinco? La respuesta es de nuevo 6 turnos. La razón es que cada tirada es completamente independiente de las anteriores. El dado no "recuerda" cuánto hace que salió un cinco, de modo que la probabilidad de que salga en el primer turno a partir de la incorporación del jugador es 1/6, la de que salga en el segundo turno es 5/6 x 1/6, y así sucesivamente. Es decir, la probabilidad de que salga el cinco en el enésimo turno es exactamente igual a la probabilidad de que la distancia entre dos cincos consecutivos sea n. Por tanto, el número medio de tiradas que nuestro jugador tiene que esperar hasta que salga el primer cinco es también 6. Esto es en cierto modo sorprendente, sobre todo si planteamos la cuestión de otro modo, tal y como se hace en el gráfico superior de la figura 3. En rojo se muestran las distintas apariciones de los cincos. La flecha verde indica el momento en el que el jugador se incorpora a la partida. De la figura parece desprenderse que la distancia entre dos cincos consecutivos será siempre mayor que el número de turnos que tiene que esperar el jugador



1. 35 tiradas de dado. El cinco ha salido 7 veces y las llaves indican el número de turnos entre un cinco y el siguiente.



2. Probabilidad de que haya que esperar *n* turnos hasta la aparición del primer cinco, en el juego del dado (*rombos azules*) y en el juego donde los números siguen una secuencia periódica (*cuadrados rojos*).

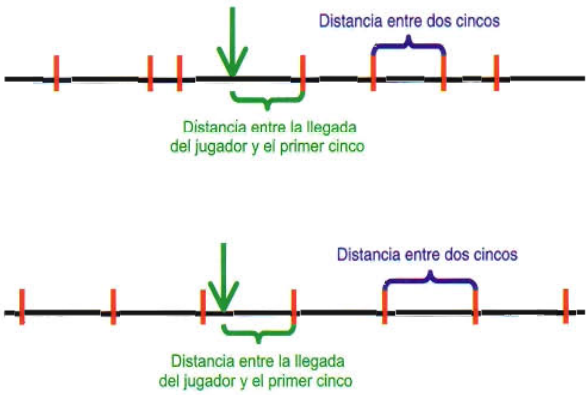
para ver el primer cinco. Sin embargo, la "falta de memoria" del dado nos asegura lo contrario. El valor medio de ambas cantidades es exactamente el mismo.

Pensemos en una variante, aparentemente trivial, del juego. Supongamos que en lugar de un dado, los números van saliendo consecutivamente 1, 2, 3, 4, 5, 6, 1, 2, 3, etc. En este caso, la distancia entre dos cincos consecutivos es siempre la misma: 6 turnos, tal y como se muestra en el gráfico inferior de la figura 3. Si un jugador se incorpora a la partida en un turno elegido al azar, el número que aparece en ese turno puede ser cualquiera de los seis y el número de turnos que tiene que esperar hasta el primer cinco puede ser 1, 2, 3, 4, 5 ó 6, todos ellos con la misma probabilidad. Por tanto, el valor medio de este número de turnos es:

$$\langle n \rangle = \frac{1}{6} \times (1+2+3+4+5+6) = 3{,}5$$

Este resultado es más acorde con nuestra intuición. Si entre dos cincos hay una cierta distancia media, lo que uno espera es que la distancia entre la llegada del jugador y el siguiente cinco sea, más o menos, la mitad. En el caso de la secuencia fija de números, esto es lo que ocurre. Sin embargo, en la secuencia de tiradas del dado no es así: la distancia media entre cincos es igual a la distancia media entre la llegada y el siguiente cinco. Este hecho se conoce como la *paradoja del autostopista*, porque puede también reformularse del siguiente modo. Un autostopista llega a una carretera muy poco concurrida, por donde los coches pasan aleatoriamente, pero de forma que el tiempo medio entre un coche y otro es de 15 minutos. El autostopista cree que tendrá que esperar unos 7 minutos y medio hasta la llegada del siguiente coche, pero no es así. Si los coches aparecen de la misma forma que los cincos en nuestro juego, entonces el tiempo medio de espera será también de 15 minutos. Si el autostopista, en lugar de esperar coches que llegan aleatoriamente, pudiera ir a una parada de autobús en donde éstos llegaran, pongamos, cada 15 minutos (y con puntualidad), entonces el tiempo medio de espera sí sería de 7 minutos y medio. El esquema de ambas situaciones es idéntico al de la figura 3.

La paradoja del autostopista se ha aplicado recientemente al transporte de partículas brownianas. Cualquier partícula pequeña en el seno de un líquido o un gas, como una proteína en el interior de la célula, una mota de hollín en el aire o un grano de polen en agua, experimenta un movimiento errático que se llama movimiento browniano, debido a que el primero en observarlo fue el botánico inglés Robert Brown, en 1827, al mirar a través del microscopio unos granos de polen inmersos en agua. Desde hace unos años, se están estudiando mecanismos para transformar este movimiento errático en un movimiento dirigido en una dirección. Uno de estos mecanismos es una especie de válvula que atrapa a la partícula browniana que llega por la izquierda, la deja pasar hacia la derecha, pero no le permite regresar. Hernán L. Martínez, de la Universidad estatal de California, y yo mismo hemos estudiado distintos modos de redistribuir estas válvulas a lo largo de una línea. Si las válvulas aparecen y desaparecen cada cierto tiempo, resulta más efectivo diseminarlas al azar que periódicamente. La razón es precisamente la paradoja del autostopista. Piensen en la línea horizontal de la figura 3 como el espacio que recorre la partícula. Las líneas verticales rojas en la figura 3, que antes representaban las apariciones del cinco o las llegadas de los coches, serían ahora las válvulas. Cuando éstas reaparecen, la partícula se encuentra en un punto al azar, que estaría representado por la flecha verde de la figura 3. La distancia que puede llegar a recorrer la partícula en su movimiento errático es precisamente la que abarcan las llaves verdes de la figura. Hemos visto que esa distancia, en media, es casi el doble en el caso de la distribución aleatoria que en el caso de la distribución periódica. Por tanto, si se distribuyen las válvulas aleatoriamente, se podrán transportar las partículas brownianas hacia la derecha más rápido que si se distribuyen periódicamente. Este es otro caso curioso en el que el desorden o el azar mejoran el rendimiento de un sistema.



3. Representación esquemática de los dos juegos descritos en el texto. En el primero se lanza un dado, mientras que en el segundo los números aparecen consecutivamente. En ambos, el jugador llega en algún turno escogido al azar.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Matemáticas sostenibles

Alicia y Bruno quieren explotar un pequeño manantial de agua mineral. La productividad del manantial, es decir, el número de litros que se pueden extraer en un día, depende del uso que se le haya dado en los días anteriores. Si un día se extrae agua del manantial (supondremos que o bien no se extrae nada o se extrae la cantidad determinada por la productividad del día), la productividad disminuye en un litro diario. Si un día no se extrae nada y se permite así que la capa freática del manantial se recupere, su productividad aumentará en un litro diario, siempre que no exceda los 100. El primer día la productividad es la máxima, es decir, 100 litros diarios.

Los dos socios discuten sobre la forma óptima de explotar el manantial. Bruno opina que lo mejor es extraer agua siempre que la productividad no sea nula, pero Alicia pronto le demuestra que ésa no es una buena idea. Es cierto que extraerán agua los cien primeros días, pero en el día 101 la productividad será nula y no podrán extraer nada. Tendrían que esperar otros 100 días sin utilizar el manantial para que éste recuperase su productividad máxima de 100 litros diarios. Alicia se inclina por la siguiente estrategia: extraer agua un día sí y otro no. De este modo la productividad oscilará entre 100 y 99 y se extraerán 100 litros cada dos días, manteniendo una productividad media de 50 litros por día.

Sin embargo, después de hacer algunas cuentas Bruno volvió a defender su idea inicial. "Imagina —dice Bruno— que explotamos el manantial durante 100 días tal y como yo propongo, extrayendo agua todos los días. El primer día extraemos 100 litros y la productividad baja a 99 litros por día. El segundo día extraemos 99 litros y la productividad baja a 98 litros/día, y así sucesivamente. En el centésimo día, extraemos sólo 1 litro y la productividad baja a cero. En los cien días, habremos extraído un total de

$$100 + 99 + 98 + \ldots + 2 + 1 = \frac{(100 + 1) \times 100}{2} = 5050 \text{ litros.}$$

Ante la expresión de extrañeza de Alicia, Bruno le replica: "¿Ya no te acuerdas cómo se suma una progresión geométrica, como 1, 2, 3,..., 98, 99, 100? La suma es igual al número de sumandos multiplicado por la media del primero y el último."

"Calculemos ahora —prosigue Bruno— la producción total utilizando tu política de extracción cada dos días. En este caso extraemos 100 litros el primer día, 100 litros el tercero, 100 el quinto, etc. Al final, habremos utilizado el manantial 50 días. Por tanto, la producción total será de $50 \times 100 = 5000$ litros, que es menor de lo conseguido con mi estrategia."

"Sí —responde Alicia—, pero a partir del centésimo día tienes el manantial seco como la mojama. Te puedo demostrar que la mejor política *sostenible* es la mía. Por sostenible entiendo que el manantial se va a utilizar por un tiempo indefinido. Imagina que extraemos agua n días seguidos, dejamos descansar el manantial otros m días seguidos, y repetimos esta secuencia continuamente. Está claro que n tiene que ser igual a m. Si fuera mayor, la productividad acabaría siendo nula y, si fuera menor, estaríamos dejando de extraer agua incluso con productividad máxima. Es decir, tenemos que extraer agua n días y dejar descansar el manantial otros n días para que recupere su productividad inicial de 100 litros diarios. Al cabo de esos $2n$ días, habremos extraído una cantidad de agua:

$$100 + 99 + 98 + \ldots + (100 - n + 1) = \frac{(201 - n)\, n}{2} \text{ litros.}$$

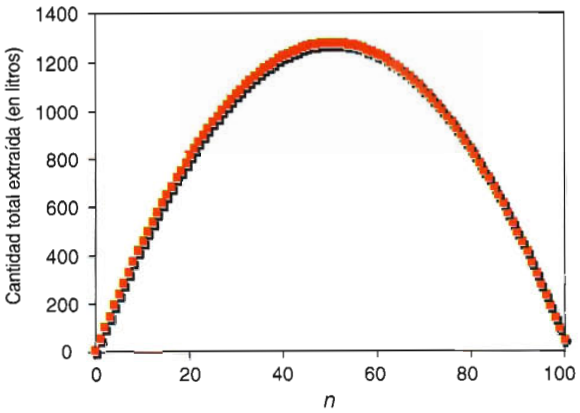
La productividad media se obtiene dividiendo esta cantidad total de agua por el número de días; es decir:

$$\frac{(201 - n)\, n}{2 \times (2n)} = \frac{201 - n}{4} \text{ litros/día;}$$

y es mayor cuanto menor sea n. El valor óptimo es por tanto $n = 1$, que coincide con la estrategia que yo defendí desde el comienzo: sacar agua un día sí y otro no y mantener la productividad siempre entre 100 y 99 litros por día".

"Acepto la derrota —dijo Bruno. Aun así, mi estrategia es mejor que la tuya si tuviéramos que explotar el manantial sólo durante 100 días. ¿Cuál es entonces la mejor estrategia si el plazo de explotación es de N días?"

La pregunta de Bruno no es sencilla. Si el plazo es muy amplio, es decir, si N es muy grande, entonces parece claro que lo más conveniente es adoptar la estrategia de Alicia durante el comienzo de la explotación. Sin embargo, cuando el final del plazo se acerca conviene extraer más agua a pesar de disminuir la productividad del manantial, puesto que no nos preocupa cómo quede éste cuando el plazo de explotación venza. Analicemos entonces la siguiente estrategia: durante $N - n$ días utilizaremos la estrategia sostenible y en el resto, es decir, en los últimos n días, utilizaremos el manantial a diario. Llamaré *período sostenible* a los primeros $N - n$ días y *período de extracción exhaustiva* a los últimos n días. También tomaremos n de forma que $N - n$ sea par. De este modo, el último día del período sostenible no se extrae agua (recordemos que la estrategia sostenible es extraer el primer día, descansar el segundo, y así sucesivamente)



1. Cantidad total de agua extraída (obviando el término $50N$) en función del número de días n del plazo de extracción exhaustiva.

y el período de extracción exhaustiva comenzará con productividad 100 litros/día. El agua extraída en el período sostenible es

$$\frac{N-n}{2} \times 100 \text{ litros}$$

y la producción total en el período de extracción exhaustiva es

$$100 + 99 + 98 + \ldots + (100 - n + 1) = \frac{(201 - n)\,n}{2} \text{ litros.}$$

Por lo tanto, la cantidad total extraída con nuestra estrategia combinada es

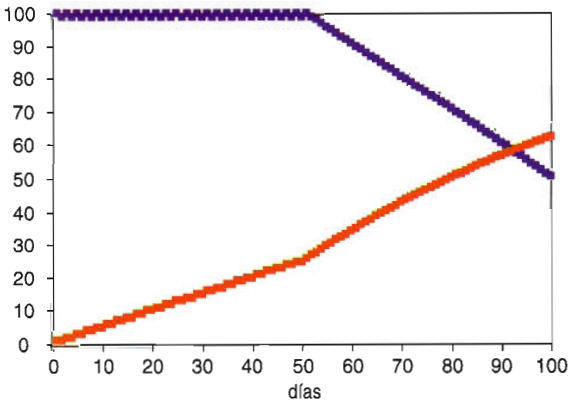
$$\frac{100\,N + 101n - n^2}{2} \text{ litros.}$$

Debemos encontrar el valor de n que hace máxima esta cantidad. En la figura 1 se muestra la cantidad total extraída (obviando el término proporcional a N), cuya gráfica es una parábola invertida. El vértice de la parábola está en $n = 50{,}5$ días. Encontramos por tanto que la cantidad total extraída alcanza su máximo tanto para $n = 50$ días como para $n = 51$ días. Para ambos casos, esta cantidad vale $50N + 1275$ litros. Por lo tanto, la mejor estrategia para un plazo de explotación de N días es extraer agua un día sí y otro no hasta que falten 50 (si N es par) o 51 días (si N es impar) para la finalización del plazo. A partir de ese momento lo que hay que hacer es extraer agua todos los días. Dejamos el manantial con una productividad de 50 litros por día, la mitad de su máximo valor posible. Pero esto nos importa poco: le tocará al siguiente esperar a que el manantial se recupere.

Para el caso de $N = 100$ días, el período sostenible durará 50 días y el de extracción exhaustiva otros 50. La cantidad total extraída es de 6275 litros, bastante superior a los 5000 de la estrategia sostenible y a los 5050 de la estrategia que proponía Bruno. En la figura 2 he representado la productividad y la producción total en función del tiempo. Vemos que la productividad oscila entre 100 y 99 durante el período sostenible y luego baja hasta 50 litros por día.

Pero, ¿es ésta la mejor estrategia posible? ¿No puede haber otras combinaciones con mayor producción? ¿Combinaciones en las que la productividad, por ejemplo, baje lentamente y no tenga el comportamiento abrupto de la figura 2? Yo no he sido capaz de encontrar ninguna, pero emplazo a los lectores a hacerlo o, en su defecto, a encontrar un argumento que pruebe que la estrategia de la figura 2 es la óptima.

Luis Dinís, de la Universidad Complutense de Madrid, Chris van den Broeck y Bart Cleuren, del Limburg Universitair Centrum (Bélgica), y yo mismo, hemos trabajado en estos problemas de optimización. Van den Broeck llama "estrategia avariciosa" a la de Bruno, puesto que trata de obtener el máximo posible cada día sin preocuparse de lo que ocurrirá al día siguiente. Dinís y yo también solemos referirnos a ella como "matar a la gallina de los huevos de oro". Ambas denominaciones indican que la estrategia de Bruno es una optimización de las ganancias inmediatas que resulta muy negativa a largo plazo, algo que desgraciadamente ocurre con frecuencia en la gestión política y empresarial de nuestro tiempo.



2. Productividad (*puntos azules*), en litros por día, y producción acumulada (*puntos rojos*), en miles de litros, en función del número de días transcurridos. En el día 50 se produce el cambio de la estrategia sostenible a la de extracción exhaustiva y la productividad comienza a bajar.

La combinación de una estrategia sostenible con un período de explotación exhaustiva puede darse en otras situaciones. Un atleta de medio fondo, por ejemplo, corre a una velocidad estacionaria la mayor parte de la carrera y realiza un *sprint* en los últimos metros, en donde realmente agota hasta la última de sus fuerzas. Un amigo ex atleta me informó de que un corredor emplea más de dos ritmos en una carrera debido a la presencia de los otros corredores. Sin embargo, en una contrarreloj, me aseguró, utilizaría sólo dos: el régimen "estacionario" y el *sprint*, como en la estrategia de la figura 2. Para el manantial, los últimos 50 días de explotación serían el equivalente al *sprint*: una explotación intensiva que acaba dejándolo medio exhausto.

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

El reparto del poder en la Unión Europea

En los últimos meses ha habido una intensa discusión política acerca del reparto de poder en la futura Constitución de la Unión Europea. Uno de los puntos más polémicos es la regla por la que el Consejo de Ministros de la Unión Europea aprueba o rechaza proposiciones de la Comisión y el Parlamento Europeo. Esta regla ha variado considerablemente en el último año, desde el Tratado de Niza hasta el borrador final de la Constitución Europea aprobado el pasado mes de junio.

El Consejo está formado por los primeros ministros de los 25 países integrantes. En el tratado de Niza se asignaba un peso al voto de cada estado miembro, mientras que el borrador de la Constitución Europea establece el sistema de la doble mayoría. Cualquier resolución del consejo necesita los votos de al menos 15 estados miembros y que la población total de los estados que votan afirmativamente alcance un 65 % de la población total de la Unión. En el pasado mes de junio se relajó este segundo requisito: para que una proposición sea denegada es necesario que se opongan a ella al menos cuatro países. De modo que si se oponen sólo tres países, la proposición sería aceptada aun sin alcanzar el 65 % de la población. Esta última disposición impide esencialmente que tres estados grandes, como Alemania, Francia y Reino Unido, por ejemplo, puedan bloquear una proposición en contra del resto del Consejo.

La discusión en nuestro país se ha centrado en la pérdida de poder que supone la Constitución en relación con el Tratado de Niza. ¿Existe algún método para medir el poder de una nación en el Consejo?

El problema del "poder" de un individuo o país en un sistema de votación ponderado ha sido estudiado matemáticamente desde 1940. Pronto los investigadores se dieron cuenta de que el "poder" no es proporcional al número de votos. Supongamos una empresa con dos accionistas. Uno de ellos tiene el 40 % de las acciones y, por tanto, tiene 4 votos en el consejo de accionistas, mientras que el otro posee el 60 % y dispone de 6 votos. Es evidente que el accionista mayoritario tiene todo el poder si las resoluciones del consejo se toman por mayoría simple, mientras que el poder del accionista minoritario es nulo: sus votos, a pesar de constituir un 40 % de los votos totales, no tienen ninguna influencia en el consejo. Lionel Penrose (padre de los conocidos físicos Roger y Oliver Penrose) en Inglaterra y John F. Banzhaf en los Estados Unidos estudiaron el problema y llegaron a la misma solución de modo independiente. Definieron una cantidad, que hoy se conoce como índice de poder de Banzhaf, que mide el poder real de un votante en un consejo.

El índice de poder de Banzhaf tiene en cuenta las coaliciones que pueden formar los votantes del consejo. Si hay N votantes, el número de coaliciones que pueden formar (incluyendo la totalidad del consejo y la coalición vacía) es 2^N. De ellas, algunas serán ganadoras, es decir, capaces de sacar adelante una resolución, y otras perdedoras. Decimos que un votante es *indispensable* en una coalición ganadora si, al abandonarla, ésta se convierte en perdedora. Es evidente que el poder de un votante es alto si es indispensable en muchas coaliciones. El índice de poder de Banzhaf de un votante x es proporcional al número ω_x de coaliciones en las que el votante es indispensable. Más precisamente, el índice es igual a la probabilidad de que el votante sea indispensable en una coalición elegida al azar de entre todas las que el votante puede formar con sus compañeros. Como hay 2^{N-1} coaliciones que contienen a x, el índice de poder de Banzhaf será

$$\eta_x = \omega_x / 2^{N-1}.$$

Veamos un ejemplo concreto. Supongamos un consejo integrado por tres personas, A, B y C, que tienen, respectivamente, 2, 5 y 6 votos, y en el que son necesarios 8 votos para aprobar una resolución. Es decir, la resolución se aprueba sólo si obtiene 8 o más votos. Este número mínimo de votos necesarios para aprobar la resolución se llama *cuota*. Las coaliciones ganadoras son la A, B, C, es decir, la unanimidad, la A, C y la B, C. El votante A es indispensable en la segunda coalición, B es indispensable en la tercera y C lo es en todas ellas. Por tanto, el índice de poder del votante A es $\eta_A = 1/2^2 = 1/4$, el índice de B es también $\eta_B = 1/4$, mientras que el de C es $\eta_C = 3/2^2 = 3/4$. Vemos que los votantes A y B tienen el mismo poder, a pesar de que B tiene más del doble de votos. Sin embargo, C, con sólo un voto más que B, tiene tres veces más poder.

En ocasiones, en lugar de dividir el número de coaliciones decisivas de un votante entre el número total 2^{N-1} de coaliciones que lo incluyen, se toma el tanto por ciento de coaliciones decisivas del votante x en relación a la suma total de coaliciones decisivas para todos los votantes. Este índice se llama índice de poder de Banzhaf *normalizado*, mientras que el que hemos definido en el párrafo anterior se denomina *probabilístico*. En nuestro ejemplo, la siguiente tabla muestra las coaliciones decisivas, el índice de Banzhaf probabilístico y el normalizado:

VOTANTE	NUMERO DE VOTOS	COALICIONES DECISIVAS	INDICE DE BANZHAF NORMALIZADO	INDICE DE BANZHAF PROBABILISTICO
A	2	1	20%	1/4
B	5	1	20%	1/4
C	6	3	60%	3/4

Los dos índices guardan la misma proporción entre sí. Por ejemplo el de C es en los dos casos, probabilístico y normalizado, tres veces el de A y B. Los índices normalizados, aunque no tienen ninguna interpretación probabilística, suman siempre 100, por lo que son más útiles cuando se trata de comparar índices de poder entre dos situaciones distintas, como la Constitución Europea y el Tratado de Niza, por ejemplo.

Jesús Mario Bilbao, catedrático de matemática aplicada de la Universidad de Sevilla y cuyos comentarios y sugerencias han sido indispensables para la elaboración de este artículo, ha calculado estos índices para los distintos sistemas de votación propuestos para el Consejo de Ministros de la Unión Europea, desde el ya obsoleto Tratado de Niza hasta las últimas versiones de la Constitución. Los resultados completos pueden verse en su página web http://www.esi2.us.es/~mbilbao. Según los cálculos del profesor Bilbao, el índice de poder normalizado de los seis países más poblados de la Unión es:

EU 25	POBLACION	TRATADO DE NIZA	13 estados y 60% de la población	15 estados, 65% de la población y bloqueo de más de 4 naciones
ALEMANIA	18,158	8,5606	13,360	10,424
FRANCIA	13,118	8,5600	9,4887	7,5805
REINO UNIDO	18,052	8,5600	9,4281	7,5395
ITALIA	12,610	8,5600	9,1807	7,3818
ESPAÑA	9,141	8,1221	7,0202	5,8233
POLONIA	8,408	8,1221	6,7677	5,5566

En la segunda columna se detalla el porcentaje de población del país con respecto a la población total de la Unión; en la tercera el índice de poder según las normas del Tratado de Niza; en la cuarta según el primer borrador de la Constitución Europea, que establece que una proposición se aprueba con una mayoría de estados (13) que comprendan un 60 % de la población; y en la quinta el índice de poder según la versión de la Constitución acordada el pasado mes de junio, que establece que una resolución se aprueba si al menos 15 estados votan a favor y esos estados tienen un 65 % de la población, siempre que sean más de cuatro estados los que votan en contra.

El problema que suscitan estas cifras no es tanto si se pierde o gana con respecto al Tratado de Niza, sino cuál es la distribución de poder más justa. Con las reglas adoptadas en los distintos borradores de la Constitución Europea, el índice de poder es aproximadamente proporcional a la población de cada país, lo cual parece a primera vista razonable. Sin embargo, existe un argumento debido a Lionel Penrose según el cual el índice de poder debería ser proporcional a la raíz cuadrada de la población. El argumento es el siguiente. Calculemos el índice de poder de Banzhaf de un ciudadano que vota en un país con N habitantes, con N par y siendo la cuota $N/2$, es decir, la mayoría simple. Este ciudadano será indispensable sólo en las coaliciones en las que él esté y que tengan exactamente $N/2$ individuos. El número de coaliciones decisivas es el número combinatorio

$$\omega_x = \binom{N-1}{N/2} = \frac{(N-1)!}{(N/2)!(N/2-1)!}\,.$$

Con algo de matemáticas superiores, se puede demostrar que este número es aproximadamente igual a $2^{N-1}\sqrt{2/\pi N}$. Por tanto, el índice de poder del ciudadano es inversamente proporcional a la raíz cuadrada de la población del país. Si los habitantes del país votan una resolución y luego su primer ministro transmite la decisión al consejo, puede también demostrarse que el índice de poder del ciudadano en este proceso de dos niveles es el producto de los dos índices de poder: el que tiene el ciudadano dentro de su país y el que tiene su primer ministro en el Consejo. Por lo tanto, para que todos los ciudadanos de la Unión tuvieran el mismo índice de poder, el índice del primer ministro en el consejo debería ser proporcional a la raíz cuadrada de la población del país. De este modo el producto del índice del ciudadano y el del primer ministro sería el mismo para todos los países.

Inspirados por este argumento, los matemáticos polacos Slomczynki y Zyczkowski han propuesto la llamada *regla Penrose-62*, que consiste en dar a cada país un voto proporcional a la raíz de su población y establecer una cuota del 62 % de los votos totales. Una carta a los gobiernos de la Unión Europea, firmada por más de 40 investigadores de toda Europa, defiende la adopción de la regla Penrose-62 en la futura Constitución. Con esta regla los índices de Banzhaf serían también aproximadamente proporcionales a la raíz cuadrada de la población.

El problema es suficientemente serio como para que todos estos argumentos se tomen en consideración. Se puede entender que aparezcan ciertas reticencias a la adopción de una regla basada en la raíz cuadrada de la población. Un país con 4 millones de habitantes tendría 2000 votos, pero, si se divide dos mitades, cada nuevo país tendría 1414 votos, haciendo un total de 2828. Parece extraño que el país gane poder por el mero hecho de dividirse en dos. Sin embargo, hay que tener en cuenta que la voluntad de cada ciudadano se transmite al Consejo de forma diferente antes y después de la división.

En mi opinión, el cálculo del índice de poder de Banzhaf tiene algunas limitaciones. Por ejemplo, todas las coaliciones se suponen igual de probables. Sin embargo, es evidente que hay países con mayores afinidades o intereses comunes, lo que hace que ciertas coaliciones puedan surgir de manera más sencilla. En particular, el profesor Bilbao demuestra en sus trabajos que la regla que impone un mínimo de cuatro países para bloquear una resolución apenas influye en el índice de poder de los países. Sin embargo, si se tienen en cuenta las afinidades entre los países de la Unión Europea, el requerimiento de los cuatro países de bloqueo puede tener relevancia.

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

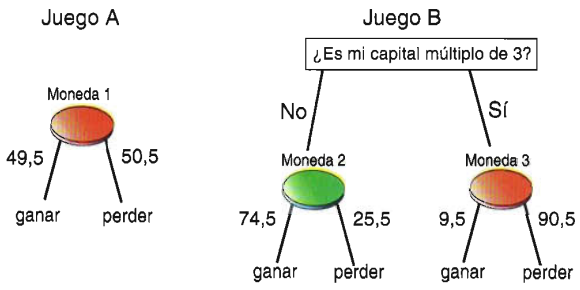
Juan M.R. Parrondo

Democracia ineficiente

Al lector asiduo de *Juegos matemáticos* probablemente le sea familiar la llamada *paradoja de Parrondo*, en la que dos juegos de azar perdedores se convierten en un juego ganador cuando se alternan periódica o aleatoriamente. Presenté la paradoja en julio de 2001 (*Perder + perder = ganar. Juegos paradójicos*) y volvimos a encontrarnos con ella en febrero de 2004 (*Las ventajas de la solidaridad*), cuando expliqué la variante debida a Raúl Toral, de la Universidad de las Islas Baleares, en la que un juego perdedor se convierte en ganador si los jugadores comparten sus ganancias entre turno y turno.

La estructura de estos juegos, a pesar de ser extremadamente simple, ha permitido el diseño de algunos modelos que presentan un comportamiento sorprendente. Luis Dinís y yo mismo hemos investigado en la Universidad Complutense de Madrid una variante de la paradoja original que muestra que, en ocasiones, la democracia puede no ser el "menos malo" de todos los sistemas de decisión. En efecto, en el modelo que hemos desarrollado, acatar la decisión de la mayoría puede dar lugar a pérdidas constantes en toda la población.

Recordemos los dos juegos de la paradoja original. Se trata de juegos en los que un individuo juega contra un casino y tiene ciertas probabilidades de ganar y perder un euro en cada turno. El primero de ellos —lo llamaremos juego A— es similar a apostar un euro a rojo o negro en la ruleta de un casino: ganamos 1€ con una probabilidad ligeramente inferior al 50 % y perdemos 1€ con una probabilidad ligeramente superior al 50 %, ya que con el cero gana siempre la banca. Supongamos que se gana con una probabilidad del 49,5 % y se pierde con una probabilidad del 50,5 % (estas probabilidades no coinciden exactamente con las de la ruleta pero nos sirven para describir la paradoja). Podemos también imaginar este juego como una apuesta sobre el resultado del lanzamiento de una moneda ligeramente sesgada. El segundo juego —lo llamaremos juego B— es un poco más complicado y, obviamente, no puede encontrarse en ningún casino. Igual que en el juego A, en cada turno ganamos o perdemos 1€ pero ahora las probabilidades dependen de lo que llevamos ganado hasta el momento (que puede ser una cantidad negativa): si lo que llevamos ganado —lo llamaremos *el capital*— es múltiplo de tres, entonces ganamos 1€ con probabilidad 9,5 %; si el capital no es múltiplo de tres, la probabilidad de ganar es del 74,5 %. Es decir, en el juego B se utilizan dos monedas, una muy desfavorable para el jugador, que se lanza cuando el capital es múltiplo de tres, y la otra bastante favorable, que se lanza cuando el capital no es múltiplo de tres. En la figura 1 se esquematizan las reglas de los dos juegos, representándose en rojo las monedas desfavorables y en verde la favorable.

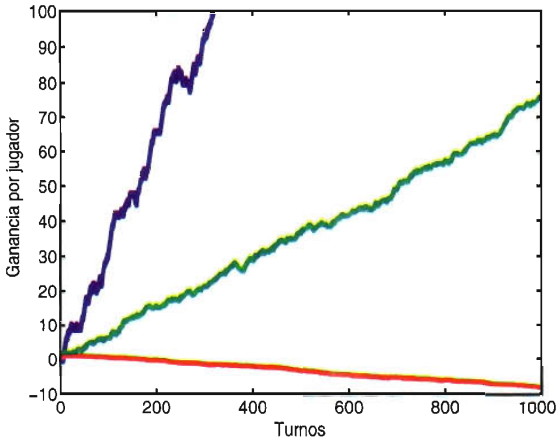


1. Las reglas de los juegos paradójicos A y B. Las monedas que se utilizan en cada juego son rojas o verdes según sean desfavorables o beneficiosas para el jugador.

Aunque se trate de juegos de azar, el jugador pierde en media si juega muchos turnos seguidos a cualquiera de los dos juegos, A o B. Sin embargo, si juega alternándolos, ya sea al azar o siguiendo alguna secuencia periódica como AABBAABB..., entonces gana en media. Esta es la paradoja original y demuestra que la alternancia de dos reglas puede dar lugar a un comportamiento muy distinto del que da cada regla por separado. El lector interesado puede consultar el artículo que dedicamos a este tema en julio de 2001; también encontrará en Internet numerosos sitios dedicados a la paradoja.

En la variante que estudio junto con Luis Dinís, son N individuos los que juegan contra el casino. En cada turno todos juegan al mismo juego pero se les permite elegir cuál de los dos juegos se va a jugar en cada turno. Si se trata de un solo jugador ($N = 1$), la elección es trivial: cuando el capital es múltiplo de tres, el jugador debe elegir A para evitar la utilización de la moneda mala de B; si el capital no es múltiplo de tres, la elección adecuada es B, puesto que entonces se utiliza una moneda muy ventajosa. Es fácil demostrar que, con esta elección, el jugador gana en media aproximadamente 1/3 de euro en cada turno.

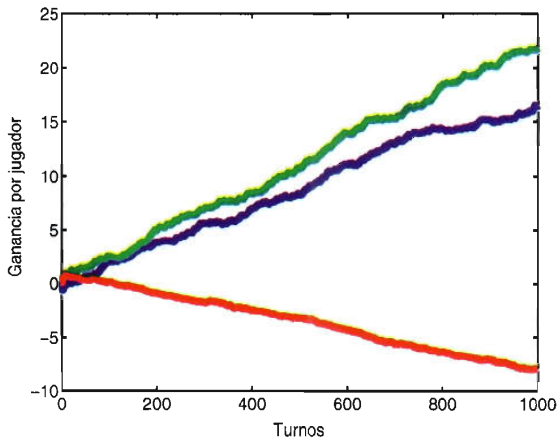
Pero, ¿qué ocurre si hay más de un jugador? En este caso, algunos de los jugadores pueden tener capital múltiplo de tres y otros no. Tienen que tomar una decisión colectiva, pero sus "intereses" son ahora opuestos. Un modo razonable de tomar la decisión es el voto. Cada jugador vota por uno u otro juego y se adopta la decisión de la mayoría. Veamos qué ocurre si se adopta esta solución democrática. En la figura 2 se muestra el capital por jugador (capital total dividido por N, el número de jugadores) en función del



2. Capital por jugador en función del número de turnos cuando el juego se elige democráticamente. La curva azul es el caso $N = 1$, la verde $N = 10$ y la roja $N = 1000$.

número de turnos y para varios valores de N. La curva azul representa la ganancia para un solo jugador $N = 1$. Vemos que, a pesar de las fluctuaciones propias de un juego de azar, la tendencia es claramente ascendente. La curva verde es la ganancia media por jugador cuando $N = 10$ jugadores y la curva roja cuando $N = 1000$. La ganancia para 10 jugadores se reduce considerablemente, comparada con el caso $N = 1$, y para 1000 jugadores el resultado es claramente desalentador: el capital disminuye lenta pero implacablemente.

En estas figuras hemos supuesto que cada jugador vota "egoístamente", es decir, vota por el juego que más le conviene en cada momento. Si su capital es un múltiplo de tres, votará por A, mientras que si el capital no es múltiplo de tres, votará por B. Se podría



3. Capital por jugador en función del número de turnos y para $N = 1000$ jugadores. La curva azul corresponde al caso en el que los juegos se eligen al azar, la roja a la elección democrática y la verde a una elección democrática en la que el 70 % de la población vota egoístamente y el 30 % restante aleatoriamente.

pensar que los 1000 jugadores no pueden hacer nada mejor que tomar esta decisión democrática. Al fin y al cabo, con ella conseguimos que la mayoría de la población juegue al juego más ventajoso. Sin embargo, el aspecto paradójico de los juegos conduce a una conclusión distinta. Recordemos que, según la paradoja original, basta con alternar los juegos, incluso de forma aleatoria, para que el resultado sea ganador. Si nuestros 1000 jugadores, en lugar de votar en función de sus intereses, eligen completamente al azar el juego que se juega en cada turno, ¡entonces ganarán! Podemos verlo en la figura 3. La curva roja muestra, de nuevo, la ganancia por jugador obtenida con la votación "egoísta", mientras que la curva azul muestra la ganancia cuando el juego se elige completamente al azar. Se pueden obtener incluso ganancias mayores si se combina una estrategia egoísta y una aleatoria. La curva verde, por ejemplo, es la ganancia cuando el juego se elige por votación pero sólo un 70 % de los jugadores vota egoístamente mientras que el 30 % restante decide su voto al azar.

¿Por qué un voto aleatorio es mejor que un voto egoísta? La razón es que el voto puramente egoísta, si N es suficientemente grande, conduce a que B se elija en todos los turnos. En efecto, para que B pierda es necesario que más de la mitad de los jugadores tengan capital múltiplo de 3, puesto que éstos son los votantes de A. Si N es grande, es muy improbable que esto ocurra. En media, sólo entre un 33 % y un 38 % de los jugadores tienen capital múltiplo de 3, tanto si se juega A como B o una combinación de ambas. Por lo tanto, la regla de la mayoría elige siempre B salvo fluctuaciones que son más raras cuanto mayor sea N. Como el juego B es perdedor, el resultado de la regla de la mayoría es la pérdida constante que reflejan las curvas rojas de las figuras 2 y 3. Sin embargo, si en el voto se incluye algo de aleatoriedad, en algunos turnos se jugará A haciendo que la paradoja entre en acción y el resultado final sea ganador.

Podríamos explicar esta ineficacia del voto egoísta en términos menos matemáticos. Las dos monedas del juego B se pueden considerar malas y buenas rachas para cada jugador. En la decisión democrática, todos los jugadores que están en buena racha quieren aprovecharla y fuerzan con su voto la elección de B. Pero la minoría de los jugadores que están en la mala racha se ven altamente perjudicados por esta decisión y el resultado global es negativo. De hecho, en cada turno la racha de un jugador cambia, por lo que el voto egoísta que le beneficia en un turno le perjudica en otro con un resultado también negativo para cada jugador individual. La moraleja es similar a la que obtenía Raúl Toral con los juegos que explicamos en febrero de 2004: conviene en ocasiones sacrificar las ganancias individuales y a corto plazo y echar así una mano a aquellos que están pasando una mala racha. No sólo porque es bueno para toda la colectividad, sino porque puede ser uno mismo quien en el futuro esté "en el agujero".

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Más sobre el reparto de poder

En los Juegos Matemáticos de agosto tratamos aspectos matemáticos del reparto de poder. Un problema con ciertas implicaciones en un asunto polémico como es el de las normas que la futura Constitución Europea propone para la toma de decisiones: la famosa *doble mayoría*. En aquel artículo analizamos el tema con cierto detalle, pero algunos lectores me han enviado preguntas y comentarios que merecen que volvamos a dedicarle una nueva entrega.

Recordemos que el *índice de Banzhaf* es una forma de medir el poder de un votante en un consejo que sigue un sistema de votación ponderado, es decir, un sistema en el que cada votante tiene un número de votos distinto. Se calcula a partir de las coaliciones que pueden formar los votantes del consejo. Si hay N votantes, el número de coaliciones que pueden formar (incluyendo la totalidad del consejo y la coalición vacía) es 2^N. De ellas, algunas serán ganadoras, es decir, capaces de sacar adelante una resolución, y otras son perdedoras. Decimos que un votante es *indispensable* en una coalición ganadora si, al abandonarla, ésta se convierte en perdedora. El índice de poder de Banzhaf de un votante x es proporcional al número ω_x de coaliciones en las que el votante es indispensable, que se llaman *coaliciones decisivas*. Más precisamente, el índice es igual a la probabilidad de que el votante sea indispensable en una coalición elegida al azar de entre todas las que el votante puede formar con sus compañeros. Como hay 2^{N-1} coaliciones que contienen a x, el índice de poder de Banzhaf será

$$\eta_x = \frac{\omega_x}{2^{N-1}}.$$

En aquel artículo, calculamos el índice de poder de Banzhaf de un ciudadano que vota en un país con N habitantes, suponiendo N par y que las resoluciones se aprueban por mayoría simple. Este ciudadano será indispensable sólo en las coaliciones en las que él esté y que tengan exactamente $N/2+1$ individuos. Con algo de combinatoria se puede demostrar finalmente que el índice de Banzhaf del ciudadano es inversamente proporcional a la raíz cuadrada de N.

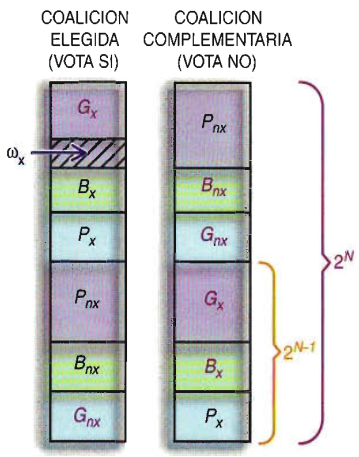
Esta dependencia con N es una de las justificaciones de la llamada norma Penrose-62, propuesta por los matemáticos polacos Slomczynki y Zyczkowski, y que consiste en dar a cada país un voto en el Consejo de Ministros de la Unión proporcional a la raíz cuadrada del número de sus habitantes y establecer una cuota (número de votos mínimos para aprobar una resolución) del 62 % de los votos totales. Con esta regla los índices de Banzhaf de todos los ciudadanos europeos serían aproximadamente iguales.

Algunos lectores han mostrado su disconformidad con este argumento, puesto que se basa en las coaliciones decisivas o coaliciones en las que el ciudadano es indispensable. Estas coaliciones deben tener exactamente $N/2+1$ individuos, lo cual las hace altamente improbables. ¿Es entonces válido un argumento que se sustenta sobre coaliciones que nunca se van a dar en la práctica?

La objeción no se sostiene porque las coaliciones decisivas son sólo un modo de presentar la definición del índice de Banzhaf. Hay sin embargo otras propiedades del índice en las que las coaliciones decisivas no son importantes. Veamos una de ellas.

Supongamos un individuo x en un consejo con N votantes. Como ya hemos mencionado, existen 2^N coaliciones posibles entre miembros del consejo. Tomemos una de ellas al azar y supongamos que la coalición elegida vota afirmativamente una resolución y el resto de los miembros del consejo votan en contra. ¿Con qué probabilidad nuestro individuo x se encontrará en el bando que gana la votación?

En la figura se esquematiza esta situación con todos los casos posibles. La coalición elegida, que vota afirmativamente, puede ser de varios tipos: ganadora (G), perdedora (P) o de bloqueo (B). Una coalición ganadora es la que gana si vota afirmativamente, una coalición perdedora es la que pierde si vota tanto afirmativa como negativamente y una coalición de bloqueo es la que gana si vota negativamente y pierde si vota afirmativamente. Esta distinción es necesaria siempre que las votaciones no se ganen por mayoría simple. Por ejemplo, si en un consejo de 50 votos sólo se aprueban las resoluciones que tienen 30 o más votos afirmativos y hay dos coaliciones enfrentadas, cada una de ellas con 25 votos, la que vota afirmativamente pierde y la que vota negativamente gana. Ambas son coaliciones de bloqueo: no son capaces de sacar adelante resoluciones, pero sí son capaces de bloquearlas. Aunque las coaliciones de bloqueo ganan las votaciones cuando votan negativamente, reservamos la denominación "coalición ganadora" únicamente para aquellas que pueden sacar adelante resoluciones. Son sólo estas coaliciones ganadoras las que se tienen en cuenta en el cálculo del índice de Banzhaf. Como se puede observar en la figura, si una coalición es ganadora, su complementa-



ria será perdedora; mientras que si una coalición es de bloqueo la complementaria será también de bloqueo. Finalmente, con el subíndice x denotamos las coaliciones a las que pertenece el individuo x y con nx aquellas a las que no pertenece.

De acuerdo con la figura, nuestro individuo x está en la coalición que gana la votación en los tres casos siguientes: cuando está en la coalición elegida y ésta es ganadora, es decir, cuando la coalición elegida es G_x; cuando no está en la coalición elegida y ésta es perdedora, es decir, cuando la elegida es P_{nx}; y, finalmente, cuando no está en la coalición elegida y ésta es de bloqueo, es decir, cuando la elegida es B_{nx}. Por tanto, la probabilidad de estar en el bando ganador es:

$$P_{ganar} = \frac{G_x + P_{nx} + B_{nx}}{2^N} = \frac{G_x + 2^{N-1} - G_{nx}}{2^N}$$

en donde hemos utilizado que $P_{nx} + B_{nx} + G_{nx} = 2^{N-1}$, algo que puede comprobarse fácilmente en la figura. Podemos ahora relacionar estos números con el número de coaliciones decisivas, ω_x. Para ello, tomemos una de las G_x coaliciones ganadoras que contienen a x. Si x se retira de la coalición, hay dos posibilidades: a) que la coalición siga siendo ganadora, con lo que, después de retirarse x, se convertiría en una de las G_{nx} coaliciones ganadoras que no contienen a x; o b) que deje de ser ganadora, con lo que la coalición de partida sería decisiva. De aquí se deduce que $G_x = \omega_x + G_{nx}$. Introduciendo esta relación en la fórmula anterior y recordando la definición del índice de poder de Banzhaf, obtenemos:

$$P_{ganar} = \frac{2^{N-1} + \omega_x}{2^N} = \frac{1 + \eta_x}{2}.$$

Esta fórmula relaciona el índice de poder de Banzhaf con la probabilidad de estar en el bando que gana la votación. Nos proporciona, por tanto, un nuevo significado del índice, en el que las coaliciones decisivas no desempeñan ningún papel.

Veamos ahora qué ocurre en una votación de dos niveles, es decir, una votación en la que un consejo está compuesto por representantes de distintos países. Si un ciudadano x del país p tiene un índice de poder η_x dentro de su país y el país tiene un índice η_p en el consejo, y suponiendo que el representante acata la decisión de la mayoría de los ciudadanos de su país, ¿cuál es la probabilidad de que dicho ciudadano esté en el bando ganador de una votación en el consejo, tomando al azar coaliciones dentro de los países? Se pueden dar dos casos: a) que el ciudadano esté en el bando ganador de su país y su representante esté en el bando ganador del consejo, o b) que el ciudadano esté en el bando perdedor de su país y su representante esté en el bando perdedor del consejo. De acuerdo con lo que hemos deducido previamente, la posibilidad a) ocurre con una probabilidad:

$$\frac{1 + \eta_x}{2} \times \frac{1 + \eta_p}{2}.$$

Por otro lado, la posibilidad b) ocurre con probabilidad:

$$\frac{1 - \eta_x}{2} \times \frac{1 - \eta_p}{2}.$$

Sumando ambas probabilidades obtenemos la probabilidad de estar en el bando ganador, que resulta ser:

$$P_{ganar} = \frac{1 + \eta_x \eta_p}{2}$$

es decir, recuperamos la misma fórmula que obtuvimos para una votación de un nivel pero con un índice de poder efectivo igual al producto de los índices de poder de los dos niveles. Por esta razón, si η_x es inversamente proporcional a la raíz cuadrada de la población del país, η_p debería ser directamente proporcional a la raíz cuadrada de dicha población, que es lo que propone la norma *Penrose-62*.

Aunque la norma Penrose-62 sale airosa de la objeción de las coaliciones decisivas, existen otras críticas más sólidas. El punto más frágil del índice de poder de Banzhaf es la hipótesis de que todas las coaliciones son igualmente probables. Esta hipótesis es equivalente a que cada votante elige su voto, afirmativo o negativo, al azar y con probabilidades respectivas del 50 %. Es una hipótesis plausible, ya que, al no tener ninguna información acerca de las posturas de los votantes, la equiprobabilidad parece ser la única elección posible. El problema es que el cálculo del índice de Banzhaf es extremadamente sensible a la probabilidad con la que cada individuo elige su voto. Por ejemplo, si en un país los ciudadanos votan afirmativamente con una probabilidad del 52 % y negativamente con el 48 %, el índice de poder de cada uno de ellos resulta ser inversamente proporcional al número de habitantes y no a su raíz cuadrada. Gelman, Katz y Bafumi, matemáticos de la Universidad de Columbia y del Instituto Tecnológico de California, han realizado un cuidadoso estudio acerca del poder real de los votantes en distintas elecciones de los Estados Unidos y concluyen que el índice de poder real está más próximo al inverso de la población que al inverso de su raíz cuadrada. De acuerdo con estas investigaciones, las formas de representación habituales, que dan un número de votos proporcional al número de habitantes, son más justas que la norma Penrose-62.

Como ven, no hay unanimidad entre los investigadores acerca de los modos de representación más justos. Es un problema que aún sigue abierto. El proyecto de Constitución Europea basa el sistema de decisión en la *doble mayoría*, que, en lo que respecta al poder de cada país en el Consejo, es similar al modo de representación tradicional, proporcional a la población y no a su raíz cuadrada. En mi opinión, es más prudente esta postura que la aplicación de una norma novedosa e interesante como la Penrose-62, pero que aún no cuenta con una total aceptación en la comunidad científica. Finalmente, conviene también recordar que el sistema de la doble mayoría tiene un importante significado político que difícilmente pueden recoger los formalismos matemáticos: la idea de que los dos pilares sobre los que se construye la Unión Europea son los ciudadanos y los estados.

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Numerogooglia

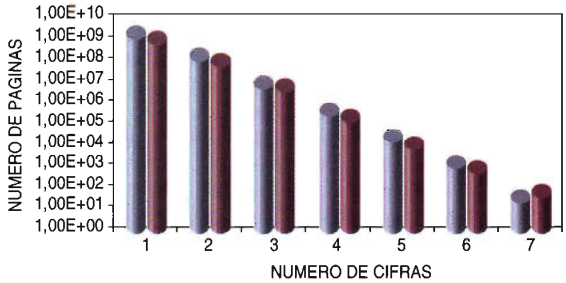
En uno de esos momentos en los que, delante del ordenador, uno no encuentra inspiración por ninguna parte, escribí casi mecánicamente mi número de teléfono en el casillero de Google, el más conocido y eficaz buscador en Internet. Google respondió inmediatamente que había encontrado 31 páginas web en las que aparecía mi número, el 3686105. El resultado me pareció sorprendente. ¿Era mi número de teléfono especial, como para estar contenido en tantas páginas, o estas apariciones eran únicamente fruto de la casualidad?

Google tiene indexadas aproximadamente 6000 millones de páginas. Suponiendo que en todas ellas hubiera un número (y sólo uno) al azar entre 1 y 10.000.000, la probabilidad de que mi teléfono estuviera en una página sería de 10^{-7} y el número medio de páginas con mi número sería $6 \times 10^9 \times 10^{-7} = 600$. Este es, por supuesto, un argumento muy grosero: ni todas las páginas contienen un solo número ni todos los números son igual de probables.

¿Podríamos refinar un poco más el argumento? En primer lugar, ¿cuántas páginas web contienen números? Una forma de estimar la fracción de páginas que contienen números, utilizando sólo las funciones de búsqueda de Google, podría ser la siguiente. Buscamos una palabra muy común, como 'the', que se encuentra en 5790 millones de páginas. A continuación buscamos la misma palabra pero eliminando las páginas que contienen '1', '2', '3', etc. Lamentablemente, Google sólo permite un máximo de 10 palabras en sus búsquedas. Por ello, sólo podemos buscar páginas que contengan 'the' (hay que poner un + antes de la palabra para que el buscador la admita) y que no contengan nueve cifras. Haciendo distintas pruebas, vemos que el número de páginas que quedan después de eliminar las que contienen números es, aproximadamente, de 26,6 millones. Esto significa que el 99,54 % de las páginas contienen las primeras cifras. Sin embargo, este método no es fiable y el resultado varía si en lugar de 'the' se utilizan otras palabras.

Por alguna razón que desconozco, el algoritmo de Google para eliminar palabras no conduce a resultados consistentes. Por ejemplo, si buscamos 'time' obtenemos 405 millones de páginas. Buscando 'time –space', es decir, las páginas que contienen la palabra 'time' y que no contienen la palabra 'space', obtenemos 24,5 millones. Finalmente, buscando 'time space', es decir, todas las páginas que contienen las dos palabras, lo que se obtiene es 10,7 millones. Evidentemente, el número de páginas con 'time' y 'space', 10,7 millones, más el número de páginas con 'time' y sin 'space', 24,5 millones, debería ser igual al número de páginas con 'time'. Sin embargo, no llega a la décima parte. Con otros pares de palabras se obtienen resultados similares. De modo que utilizar la opción de eliminar palabras no es nada fiable.

En la segunda parte del argumento inicial, suponíamos que todos los números de menos de siete cifras aparecen por igual. Esta suposición es incorrecta. Los números con menos cifras son evidentemente más comunes. Lo podemos ver buscando en Google mi número de teléfono y quitándole cifras de derecha a izquierda, es decir, buscando 3, 36, 368, 3686, hasta 3.686.105. El resultado puede verse en la figura 1, en donde también he incluido el resultado del mismo experimento realizado con otro número de siete cifras tomado al azar, el 5.477.232. El comportamiento de las dos gráficas, en escala logarítmica, es similar: el logaritmo del número de páginas decrece de manera casi lineal y con una pendiente de 7/6 = 1,17. Esto



1. Número de páginas, en escala logarítmica, que contienen las primeras cifras de dos números de siete cifras tomados al azar: 3.686.105 (*barras azules*) y 5.477.232 (*barras moradas*).

quiere decir que, al añadir una cifra, el número de páginas se multiplica aproximadamente por $10^{-7/6} = 0{,}068$. Podemos pensar que este factor es el producto de dos probabilidades: la probabilidad de que en una página aparezca un número con una cifra más, que sería del 68 %, y la probabilidad de que la cifra adicional sea precisamente la de nuestro número, que tendría que estar en torno al 10 %.

Este resultado sí parece consistente y permite obtener la siguiente aproximación para el número de páginas que contienen un número dado de c cifras:

$$N = n_0\,(0{,}068)^{c-1}$$

en donde n_0 es el número de páginas que contienen la primera cifra del número. Aunque enseguida veremos un modo de refinar el cálculo de n_0, se puede comprobar, buscando páginas que contengan el 1, el 2, etc., que n_0 está en torno a los 1000 millones de páginas. Insertando este dato en la fórmula anterior, obtenemos que un número concreto de 7 cifras debería aparecer en unas 100 páginas, uno de 8 cifras en unas 7 y uno de 9 cifras sólo en 0,46 páginas. Se trata, por supuesto, de una aproximación muy grosera,

pero el lector puede comprobar alguna de estas predicciones. Por ejemplo, pruebe a introducir números de ocho cifras y verá que Google encuentra, como mucho, una o dos páginas. Quítele una cifra al número y verá que se encuentran unas pocas decenas, tal y como predice nuestra fórmula.

La aproximación se puede refinar un poco más teniendo en cuenta una vieja y misteriosa amiga: la ley de Benford, a la que dedicamos los *Juegos Matemáticos* de diciembre de 2002. La ley dice que, en un conjunto de datos numéricos que provengan del "mundo real", la probabilidad de que el primer dígito de los datos sea un 1 es mayor que la de que sea un 2, un 3, etc. Con más precisión, la probabilidad de que el primer dígito sea d es, según la ley de Benford:

$$P_d = log_{10}\left(1 + \frac{1}{d}\right)$$

Utilizando esta fórmula se encuentra que, para $d=1$ la probabilidad es, aproximadamente, del 30 %, para $d=2$ del 17,6 %, para $d=3$ del 12,5 %, para $d=4$ del 9,7 %, para $d=5$ del 7,9 %, para $d=6$ del 6,7 %, para $d=7$ del 5,8 %, para $d=8$ del 5,1 % y para $d=9$ del 4,6 %. Como vemos, el 1 es más de seis veces más probable que el 9, una diferencia considerable. ¿Se verifica la ley de Benford en Internet?

He realizado dos experimentos relacionados con la ley de Benford. En el primero de ellos busco en Google números de sólo una cifra: el 1, el 2, etc., y calculo las frecuencias relativas, es decir, el número de páginas en que aparece un dígito concreto dividido por el número de páginas en donde aparece uno cualquiera de los 9 dígitos. El resultado se muestra en la figura 2 (*barras moradas*) y se compara con la ley de Benford (*barras azules*). El predominio del dígito 1 no es en los datos reales tan grande como el predicho por la ley; sin embargo, la tendencia descendente es clara.

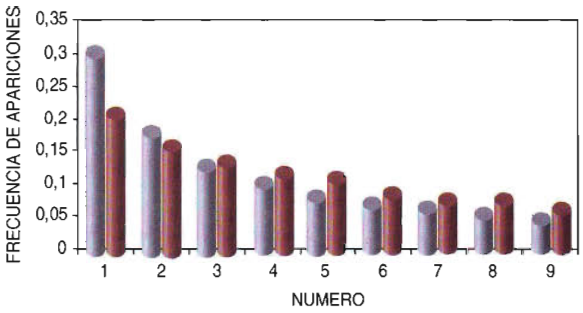
Pero la ley de Benford se aplica al primer dígito de los números y no a las veces que aparecen los dígitos por separado. Un segundo experimento más acorde con el significado de la ley es el siguiente. Tomamos un número de seis cifras al azar y vamos cambiando su primer dígito desde el 1 hasta el 9. Por ejemplo, tomando el 183.954, buscaríamos este número, luego el 283.954, el 383.954, etc. Anotamos los resultados de las búsquedas y calculamos de nuevo las frecuencias relativas. En la figura 3 se ven las frecuencias comparadas con la fórmula de Benford. El acuerdo es mayor que en el caso de la figura 2.

Estamos ahora en condiciones de refinar aún más nuestra fórmula para el número de apariciones, sustituyendo n_0 por un número proporcional a la frecuencia dada por la ley de Benford. Con ello, el número de páginas en las que debería aparecer un número de c cifras cuya primera cifra es d es:

He obtenido el factor 45×10^8 imponiendo que la

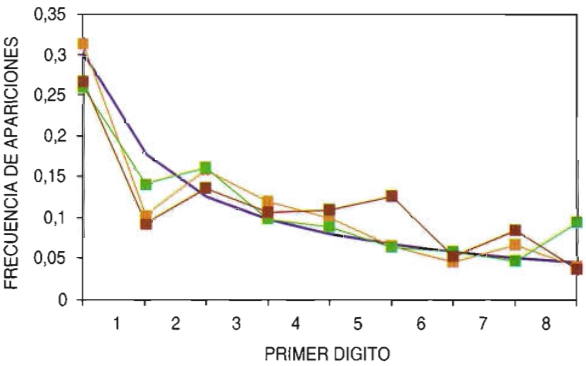
$$N = 45 \times 10^8 \times log_{10}\left(1 + \frac{1}{d}\right) \times (0{,}068)^{c-1}$$

suma de páginas para todo valor de d y de c sea de 10.000 millones, que es un poco superior a la suma de todas las apariciones de números de una cifra (evidentemente, supera el número de páginas indexadas porque en una misma página puede haber varios números). Para un número de 7 cifras comenzando por 1, la fórmula nos da 134 páginas, mientras que para uno que comienza por 3 es sólo de 56 páginas. Jugando con algunos números escritos al azar, se puede comprobar que la fórmula predice aceptablemente los resultados de las búsquedas.



2. Frecuencia con la que aparecen números de una sola cifra (*barras moradas*) comparada con la ley de Benford (*barras azules*).

Podemos ahora volver a la pregunta inicial: ¿es mi número de teléfono especialmente frecuente o sus 31 apariciones pueden ser fruto del azar? La fórmula anterior nos dice que un número de 7 cifras comenzando por 3 debería aparecer en unas 56 páginas. Por lo tanto mi número es relativamente raro. Comparen las modestas 31 apariciones con las 199.000 del número de 7 cifras 1.234.567. Este sí es un número peculiar. ¿Podría utilizarse alguna vez Google para cuantificar lo "raro" o lo "familiar" que es un número?



3. Frecuencia de apariciones cambiando el primer dígito de distintos números de seis cifras, comparada con la ley de Benford (*curva azul*).

No lo sé, pero creo que Google, además de ser una herramienta insustituible para el internauta, es un estupendo banco de pruebas para jugar con números. Invito al lector a continuar esta *numerogooglia* con sus propios experimentos o juegos. Aquí van algunas preguntas: ¿cuál es el número de 7 cifras con menos apariciones? ¿Cuál es el que aparece más veces? Curiosamente, no es el 1.234.567. He encontrado uno que aparece en más de 600 mil páginas. Imaginen cuál.

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Calculistas prodigiosos

¿Es posible calcular la raíz 13 de un número de 100 cifras en 0,15 segundos? Un calculista colombiano, Jaime García Serrano, asegura ostentar este récord en el libro Guiness. Desgraciadamente, es algo que no he podido comprobar. Tampoco he podido averiguar cómo se realizó la prueba. La raíz trece de un número de 100 cifras tiene entre 7 y 8 cifras y supongo que se tarda más de 0,15 segundos en escribirla o decirla de viva voz. Quizá los 0,15 segundos se refieren a lo que tardó García Serrano en *empezar* a dar la solución.

En cualquier caso, algo que me resultó curioso al documentarme sobre calculistas prodigiosos es que muchos de ellos se especializan en calcular raíces quintas y decimoterceras. ¿Por qué esta predilección? La respuesta es que las potencias quinta, novena y decimotercera de un número entero conservan el último dígito del número. Con ello se consigue una pequeña ventaja en el cálculo de la raíz. En esta tabla se muestra el último dígito de las distintas potencias:

ULTIMO DIGITO DE:				
n	n^2	n^3	n^4	n^5
1	1	1	1	1
2	4	8	6	2
3	9	7	1	3
4	6	4	6	4
5	5	5	5	5
6	6	6	6	6
7	9	3	1	7
8	4	2	6	8
9	1	9	1	9
0	0	0	0	0

Es una tabla fácil de construir si se tiene en cuenta que el último dígito del producto n x m depende sólo de los últimos dígitos de los dos factores. Así, para hallar el último dígito del cubo de un número que acaba en 7, basta multiplicar 7 por 9 (el último dígito de $7^2 = 49$), que es 63 y quedarse con el último dígito, es decir, 3. En la tabla vemos que el último dígito de n^5 coincide con el último dígito de n. La tabla se continuaría hacia la derecha de forma periódica, de modo que las potencias 9, 13, 17, 21, etc. de n tienen también su último dígito igual al último dígito de n. Utilizando este simple hecho matemático, el cálculo de la raíz quinta, novena o decimotercera de un número se puede simplificar bastante (siempre suponiendo que el resultado es un número entero). Con un poco de entrenamiento podemos convertirnos en calculistas prodigiosos y calcular la raíz quinta de un número de 10 cifras. Veamos cómo.

La raíz quinta de un número de diez cifras tendrá a lo sumo 2 cifras, ya que $\sqrt[5]{10^{10}} = 100$. La última de esas cifras es precisamente la última cifra del número original. Por lo tanto, sólo tenemos que calcular la primera cifra, la de las decenas. Para ello basta recordar la potencia quinta de todas las decenas, del 10 al 90, y ni siquiera es necesario recordarla con exactitud. Podemos calcular la raíz de cualquier número si memorizamos los números que aparecen en la siguiente tabla:

n	$n^5/100000$	Número a recordar
10	1	1
20	32	30
30	243	200
40	1024	1000
50	3125	3000
60	7776	8000
70	16.807	16.000
80	32.768	32.000
90	59.049	60.000

Una vez memorizada la tabla, pida a algún amigo que elija un número entre 0 y 100, que lo eleve a la quinta potencia en una calculadora y le diga el resultado. En unos pocos segundos usted podrá adivinar el número de su amigo. Supongamos que el número elegido es el 61. Su quinta potencia es 844.596.301. Sabemos entonces que el número elegido termina en 1. Para calcular sus decenas, eliminamos las cinco últimas cifras de la quinta potencia quedándonos con 8445. Este número está entre 8000 y 16.000, es decir, entre las entradas correspondientes a 60 y 70. Por lo tanto, la primera cifra del número elegido tiene que ser 6. Con un poco de práctica se puede dar la respuesta en algo menos de un segundo.

¿Habría otra forma de averiguar la cifra de las decenas u otras cifras de raíces quintas? La prueba del nueve, que aprendimos en el colegio para comprobar productos y divisiones, puede ayudarnos. Hagamos ahora una tabla con las cifras que resultan de aplicar

la prueba del nueve a las distintas potencias de un número:

PRUEBA DEL NUEVE

n	n^2	n^3	n^4	n^5	n^6	n^7
1	1	1	1	1	1	1
2	4	8	7	5	1	2
3	9	9	9	9	9	9
4	7	1	4	7	1	4
5	7	8	4	2	1	5
6	9	9	9	9	9	9
7	4	1	7	4	1	7
8	1	8	1	8	1	8
9	9	9	9	9	9	9

Para obtenerla, al igual que ocurría con la tabla del último dígito, no es necesario hacer grandes operaciones. Por ejemplo, la fila correspondiente al 5 se calcula de la siguiente forma: $5^2 = 25$ da $5 + 2 = 7$; $7 \times 5 = 35$ da $3 + 5 = 8$; $8 \times 5 = 40$ da $4 + 0 = 4$; y así sucesivamente. En este caso, salvo el 3 y el 6, cuyas potencias siempre dan 9, es en la potencia séptima cuando vuelve a aparecer el número original. Esto ocurre también con la potencia 13. Quizá por ello esta potencia sea tan utilizada por los calculistas. La potencia 13 es la más pequeña en la que el último dígito y la prueba del nueve se recuerdan con facilidad.

Con la potencia quinta, tampoco es difícil recordar el resultado de la prueba del nueve. Aparecen todos los dígitos, salvo el 3 y el 6. Lo único que hay que recordar es que el 5 y el 2 están permutados, así como el 7 y el 4. Con este nuevo truco podemos calcular de nuevo la raíz quinta de 844.596.301. Si aplicamos la regla del nueve obtenemos 4. Para ello no hace falta sumar todas las cifras, porque podemos agruparlas en pares que sumen nueve y eliminar dichos pares. Del número original, 844.596.301, podemos eliminar de un vistazo las cifras 45.963 y, con un poco más de atención, el par 8-1. Nos queda entonces 4. Según la tabla anterior, las cifras del número que buscamos deben sumar entonces 7. Como la última es un 1, la primera debe ser 6.

El método tiene por supuesto el riesgo de que la suma de las cifras de la potencia quinta sea 9, en cuyo caso la suma de las cifras del número elegido puede ser tanto 3, como 6 o 9. Sin embargo, al estar estos tres números bastante separados entre sí, podemos acudir a la tabla de las potencias quintas de las decenas, recordando ahora sólo de forma aproximada las entradas del 30, 60 y 90.

Pero puede ser mucho más espectacular combinar los dos métodos: el de la tabla de las potencias de las decenas y el de la prueba del nueve. Con ambos, se puede lograr sin mucho esfuerzo, aunque con algo de práctica, el cálculo de raíces quintas de números de 15 cifras. El número elegido tendría en este caso tres cifras. Las unidades se obtienen de forma trivial, pues coinciden con las de la potencia quinta. Las centenas se pueden calcular con una tabla parecida a la de las potencias quintas de las decenas. La tabla es idéntica a la de las decenas, con la salvedad de que en la primera columna aparecerían centenas y la segunda columna correspondería a $n/10^{10}$. Aplicaríamos el método descrito anteriormente, pero eliminando las diez últimas cifras de la potencia en lugar de las cinco últimas. Finalmente, obtendríamos las decenas con la prueba del nueve.

Por ejemplo: ¿cuál es la raíz quinta de 24.953.960.486.368? Eliminamos las últimas diez cifras y obtenemos 2495. Está entre 1000 y 3000, luego la cifra de las centenas de la raíz debe ser un 4. Por otro lado, la cifra de las unidades es 8. Finalmente, aplicando la prueba del nueve obtenemos un 1. Por tanto, la suma de las cifras de la raíz debe ser 1, con lo cual la cifra de las decenas tiene que ser necesariamente 7. El resultado es entonces 478.

Si la prueba del nueve da precisamente 9, se puede aún decidir entre las tres posibilidades. Veamos, por ejemplo, la raíz quinta de 48.524.739.602.976. La cifra de las unidades es 6. Para las centenas eliminamos los diez últimos dígitos, quedándonos con 4852, que está entre 3000 y 8000. La cifra de las centenas es entonces un 5. La prueba del nueve da 9, con lo cual la raíz buscada puede ser 516 (da 3 en la prueba del nueve), 546 (da 6) o 576 (da 9). Como 4852 está lejos tanto de 3000 como de 8000, la mejor estimación es 546, el resultado correcto. Sé que no es un método muy preciso, pero con un poco de práctica funciona en la mayoría de los casos.

Todos estos métodos muestran que, en las potencias de números enteros, hay una considerable redundancia, utilizada de manera muy hábil por los calculistas prodigiosos. Sin embargo, no hemos de restar méritos a estos magos de los números. E incluso, más allá de los trucos que utilicen, creo que son casos interesantes para indagar sobre una cuestión de bastante más calado: ¿hay una aptitud matemática intrínseca al ser humano, o son las matemáticas una pura construcción cultural? En lo que se refiere a esta cuestión, el caso más intrigante es probablemente el de Ramanujan, cuyas increíbles habilidades no se limitaban al cálculo algebraico sino que abarcaban el razonamiento abstracto en muchas ramas de la matemática.

Otro caso llamativo de calculista prodigioso, que ha aportado algo de luz al estudio de la historia de las matemáticas, es el de Thomas Fuller, un africano nacido en 1710 en la actual Liberia y que fue llevado como esclavo a los EE.UU. en 1724. Murió en Virginia a la edad de 80 años y alcanzó bastante fama como calculista. A pesar de no saber leer ni escribir, era capaz de realizar complicados cálculos con rapidez. En los últimos años, los estudiosos de este caso han llegado a la conclusión de que Fuller adquirió muchas de sus habilidades en su infancia en Africa, revelándose así que la matemática de los pueblos africanos del siglo XVIII tenía un desarrollo mayor del que se consideraba hasta ahora.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

El número mayor y la información misteriosa

¿Se puede obtener información de la nada? Hay un sencillo problema de probabilidad en el que uno tiene la impresión de que algo semejante es posible. El problema es el *juego del número mayor* y consiste en lo siguiente. Alguien elige dos números al azar, pero distintos entre sí, escribe cada uno de ellos en un papel y nos da a elegir uno de los dos papeles. Nosotros tomamos uno de los papeles y leemos su contenido. Con esta información, tenemos que adivinar cuál de los dos números es el mayor.

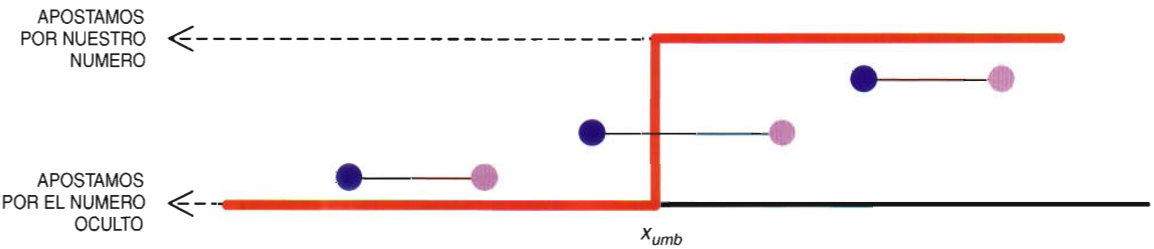
1. El juego: debemos adivinar cuál de dos números aleatorios es el mayor, conocido uno sólo de ellos.

Lo único que conocemos para adivinar cuál de los números es mayor es el número escrito en nuestro papel. No sabemos nada acerca de cómo se han elegido los números. Puede que sean positivos, negativos, entre 1 y 100 o entre 10.000 y 40.000. Parece entonces que leer el número de nuestro papel no nos puede aportar ninguna información. El otro número, el oculto, puede ser cualquiera y, por tanto, tenemos una probabilidad 1/2 de acertar, tanto si apostamos por nuestro número como si lo hacemos por el número oculto.

Sin embargo, se puede aumentar esta probabilidad con un truco sorprendentemente sencillo. Basta para ello que elijamos antes del juego un número umbral x_{umb}. Si el número que leemos en nuestro papel lo supera, diremos que dicho número es el mayor y si no lo supera, diremos que el número oculto es el mayor. ¿Es capaz esta estrategia de aumentar la probabilidad de ganar el juego? Llamemos *a* al mayor de los dos números y *b* al menor. En nuestro papel estará *a* con probabilidad 1/2 y *b* con probabilidad 1/2. Pueden darse tres casos: que tanto *a* como *b* estén por encima del umbral, que estén por debajo, o que *a* esté por encima y *b* por debajo (*véase la figura 2*). No sabemos con qué probabilidad se da cada uno de estos casos, pero lo que es seguro es que si se da el tercero de ellos, nuestro método del umbral acertará. En los otros dos casos el método del umbral tiene una probabilidad 1/2 de acertar. Por tanto, la probabilidad media de acertar es mayor que 1/2, siempre que el tercer caso ($a > x_{umb} > b$) se dé alguna vez, y es igual a 1/2 si no se da nunca. Es decir, hemos conseguido aumentar la probabilidad de ganar, siempre que coloquemos el umbral de modo tal que el caso $a > x_{umb} > b$ se dé en alguna ocasión.

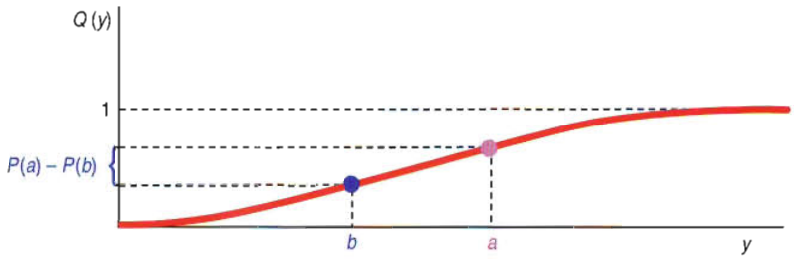
El lector puede pensar que aquí no hay nada misterioso. Podemos acertar con probabilidad mayor que 1/2, pero necesitamos algo de información para colocar el umbral adecuadamente. Si tomamos 100 como umbral y los números se eligen tirando un dado de seis caras o al azar entre 1000 y 2000, la estrategia no tiene ningún efecto. Para cuantificar mejor la influencia del umbral sobre la probabilidad de acertar, supongamos que cada uno de los dos números se elige al azar, de forma independiente, y tal que $P(x)$ es la probabilidad de que el número elegido sea menor que x. Entonces, la probabilidad del primer caso ($a, b > x_{umb}$) es $P_1 = [1 - P(x_{umb})]^2$; la probabilidad del segundo caso ($a, b < x_{umb}$) es $P_2 = P(x_{umb})^2$; y la del tercer caso ($a > x_{umb} > b$) es $P_3 = 2P(x_{umb})[1 - P(x_{umb})]$. La probabilidad de acertar es entonces:

$$P_{acertar} = \frac{1}{2}(P_1 + P_2) + P_3 = \frac{1}{2} + P(x_{umb})\,[1 - P(x_{umb})\,]$$



2. Los tres casos posibles en la estrategia del umbral: el mayor de los dos números, *a*, se representa con un punto morado y el menor, *b*, con uno azul. Cuando *a* está por encima del umbral y *b* por debajo (caso intermedio), se acierta siempre, mientras que, si ambos están por encima o por debajo (caso superior e inferior) se acierta con probabilidad 1/2.

La probabilidad de acertar más alta es $P_{acertar} = 3/4$ y se alcanza si ponemos el umbral de forma que $P(x_{umb}) = 1/2$. Por ejemplo, si los números se escogen al azar entre 1 y 100, $P(x) = x/100$ y el umbral óptimo será 50. En general, si los números se escogen al azar en un cierto rango, el umbral óptimo será el punto medio de dicho rango. Por el contrario, si $P(x_{umb}) = 0$ o 1, los dos números elegidos siempre estarán por encima o por debajo del umbral, respectivamente, y, por tanto, el método del umbral no ofrece ninguna ventaja ($P_{acertar} = 1/2$). Necesitamos por tanto conocer cómo se eligen los números para encontrar un umbral adecuado.



3. Estrategia aleatoria: si *y* es el número que encontramos en nuestro papel, apostamos por él con una probabilidad *Q*(y) (*curva roja*).

Sin embargo, una modificación del truco del umbral permite acertar con probabilidad mayor que 1/2 sin conocer nada acerca de cómo se han elegido los números inicialmente. En lugar de un umbral que nos dice por qué número apostar, elijamos una probabilidad *Q(y)* entre 0 y 1 de apostar al número *y* que hay en nuestro papel. La estrategia es en sí misma aleatoria, es decir, no nos dice por qué número apostar sino que nos da la probabilidad de hacerlo por uno u otro: si el número que leemos en nuestro papel es *y*, diremos que es el mayor con una probabilidad *Q(y)*. Lo asombroso es que esta probabilidad puede ser cualquiera, con tal de que aumente cuando lo haga *y*, como la curva roja de la figura 3. Esto es razonable: cuanto mayor sea *y* con más probabilidad diremos que es el número mayor.

Como antes, llamemos *a* al número mayor y *b* al menor. El número *y* en nuestro papel será *a* con probabilidad 1/2 y *b* con probabilidad 1/2. Acertamos si $y = a$ y decimos que es el mayor, algo que ocurre con probabilidad $Q(a)/2$; o bien, si $y = b$ y decimos que el número oculto es el mayor, algo que ocurre con probabilidad $[1 - Q(b)]/2$. Por tanto, la probabilidad de acertar es:

$$P_{acertar} = \frac{Q(a)}{2} + \frac{1 - Q(b)}{2} = \frac{1}{2} + \frac{Q(a) - Q(b)}{2}$$

que es mayor que 1/2, puesto que $a > b$ y, por tanto, $Q(a) > Q(b)$, como se muestra en la figura 3. ¡Obsérvese que el resultado es válido para cualquier par de números elegidos y para cualquier probabilidad *Q(y)*! Lo único necesario para demostrar que la probabilidad de acertar es mayor que 1/2 es que *Q(y)* crezca cuando crece *y*. Aunque parezca increíble, el argumento es correcto. Es cierto que, para encontrar la verdadera probabilidad de acertar, habría que promediar sobre los posibles valores de *a* y *b* (con *a* mayor que *b*), pero dicho promedio será siempre mayor que 1/2, por ser $Q(a) - Q(b)$ una cantidad positiva para cualquier par de valores *a*, *b* con *a* mayor que *b*. Para la elección concreta "*Q(y)* es igual a 1 cuando *y* supera x_{umb} y 0 si no lo supera" recuperamos el método inicial del umbral, y vemos que $P_{acertar}$ es 1/2 si *a* y *b* están ambos o bien por encima o por debajo del umbral, pero es 1 si *a* está por encima y *b* por debajo.

Es cierto que la ventaja sobre la elección aleatoria ($P_{acertar}$ es 1/2) puede ser mínima. Téngase en cuenta que *Q(y)* tiene que crecer de 0 a 1 cuando *y* crece de menos infinito a infinito. Por tanto, en algunos tramos *Q(y)* tiene que crecer muy despacio, como ocurre para valores pequeños o grandes de *y* en la curva roja de la figura 3. Si el rango de valores que pueden tomar los dos números elegidos está en una zona en donde *Q(y)* crece muy despacio, entonces la ventaja de nuestro método será muy pequeña. Aun así, lo sorprendente es que si *Q(y)* crece siempre, dicha ventaja será mayor que cero, admitido que los números iniciales se han obtenido al azar y de forma independiente, pero sin importar su distribución de probabilidad.

¿De dónde hemos sacado la información que nos permite predecir cuál de los dos números es el mayor con una probabilidad superior a 1/2? En mi opinión, es un auténtico misterio. Creo que la pregunta "¿cuál es el número mayor?" es extremadamente sensible a cualquier pequeña información acerca de cómo se obtienen los números. Si conocemos la distribución de probabilidad con la que se han obtenido, podemos acertar con una probabilidad 3/4, utilizando el método del umbral óptimo. Bastaría incluso conocer el rango en el cual pueden estar los números para escoger un umbral adecuado (aunque no el óptimo). La estrategia de *Q(y)* parece utilizar simplemente el hecho de que los números elegidos están en un cierto rango, o, lo que es lo mismo, ¡utiliza el simple hecho de que los números existen!

Por último, conviene recordar que las probabilidades no tienen mucho sentido cuando se aplican a un solo turno del juego. Habría que aplicar la estrategia de *Q(y)* a un gran número de turnos para observar la ventaja de la que hemos estado hablando. Pero, en el caso de jugar varios turnos, y siempre que se nos permita cambiar de estrategia en cada uno de ellos, podríamos utilizar la información de turnos anteriores para modificar *Q(y)*. Mi impresión es que la mejor estrategia sería utilizar como umbral la mediana de los números que han salido hasta el momento, es decir, un umbral que esté por debajo de la mitad de dichos números y por encima de la otra mitad. De todos modos, quizá los lectores encuentren una estrategia mejor o puedan arrojar algo de luz sobre el enigmático origen de la ventaja proporcionada por la estrategia de la figura 3.

<parr@seneca.fis.ucm.es>

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Problemas de aparcamiento

Si usted vive en una gran ciudad y comete la imprudencia de ir a un teatro del centro en su propio coche, probablemente se haya encontrado en una situación parecida a la que se muestra en la figura 1. A una cierta distancia de su destino, pongamos 500 metros, ve un lugar para aparcar. ¿Debe aprovecharlo o es mejor continuar y confiar en que va a encontrar un sitio libre más cerca del teatro al que se dirige?

La figura 1 muestra una situación muy simple en donde sólo se puede aparcar en una calle de sentido único, en cuyo extremo se encuentra el teatro. Un coche que alcance el teatro sin haber encontrado aparcamiento debe volver al comienzo por otra calle en la que no se puede aparcar e invierte un tiempo T en dar toda la vuelta. Supondremos también que cada uno de los posibles aparcamientos está ocupado con una probabilidad p, que, desgraciadamente, será cercana a 1.

La situación así descrita no es muy realista, porque normalmente uno se puede acercar o alejar del teatro por varias rutas en las que es posible aparcar. Sin embargo, la obligación de dar la vuelta a la manzana y el tiempo T que se invierte en ello constituye una forma sencilla de penalizar el intento de acercarnos al teatro despreciando sitios libres y nos va a permitir realizar algunos cálculos.

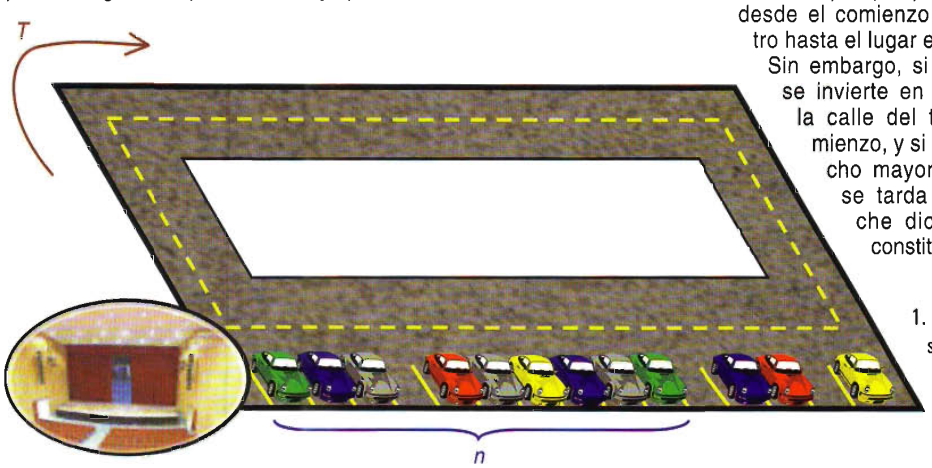
Supongamos la siguiente estrategia: aparcaremos en cualquier sitio que esté a una distancia menor o igual que n del teatro (medida en número de coches, como muestra la figura 1). ¿Cuál es el número n óptimo? Un criterio de optimización razonable es minimizar el tiempo total que tardamos en llegar al teatro. Admitamos que tardamos un segundo en atravesar andando cada posible lugar de aparcamiento y que se tardan $T = 50$ segundos en dar la vuelta en coche a todo el aparcamiento. Hay que andar un poco deprisa para ello, pero esto no afecta a nuestro problema. Se puede de hecho suponer una velocidad menor. Lo único importante es la relación entre el tiempo invertido en dar una vuelta en coche hasta el aparcamiento y el tiempo invertido en recorrer andando los sitios que nos separen del teatro.

Si empezamos a buscar aparcamiento cuando estamos a una distancia n del teatro, la probabilidad de no encontrar ningún sitio es p^n, puesto que ésta es la probabilidad de que los n sitios en donde buscamos estén ocupados. Por otro lado, encontraremos un sitio a una distancia k del teatro si los primeros $n - k$ sitios están ocupados y el siguiente libre. Esto ocurre con una probabilidad $p^{n-k}\,(1 - p)$, donde k puede ser 1, 2, 3,..., n. Por tanto, la de encontrar aparcamiento a una distancia k después de haber dado v vueltas es:

$$p^{nv}\, p^{n-k}\,(1 - p)$$

en donde v puede ser 0, 1, 2,... Se puede ver que la suma de estas probabilidades a todos los posibles valores de k y v es igual a 1, siempre que p sea menor que 1, lo cual indica que en alguna ocasión debemos encontrar sitio.

Si damos v vueltas y encontramos aparcamiento a una distancia k del teatro, ¿cuál será el tiempo total invertido en aparcar y llegar andando a nuestro destino? Como hemos supuesto que tardamos T segundos en dar la vuelta a la manzana y 1 segundo en recorrer andando cada sitio de aparcamiento, el tiempo total será $Tv + k$ segundos. En realidad, la expresión para el tiempo total es más compleja, ya que deberíamos tener en cuenta el tiempo que pasamos en el coche desde el comienzo de la calle del teatro hasta el lugar en donde aparcamos. Sin embargo, si T es el tiempo que se invierte en ir desde el final de la calle del teatro hasta su comienzo, y si este tiempo T es mucho mayor que el tiempo que se tarda en recorrer en coche dicha calle, $Tv + k$ sí constituiría una buena apro-



1. ¿Debemos aprovechar el sitio libre a una distancia n o buscar un aparcamiento más cercano?

ximación del tiempo total invertido. Supondremos que es esto lo que ocurre, aunque hay que advertir que la figura 1 no representa adecuadamente esta situación, puesto que para ello el recorrido desde el final de la calle del teatro hasta su comienzo debería ser mucho más largo que la propia calle.

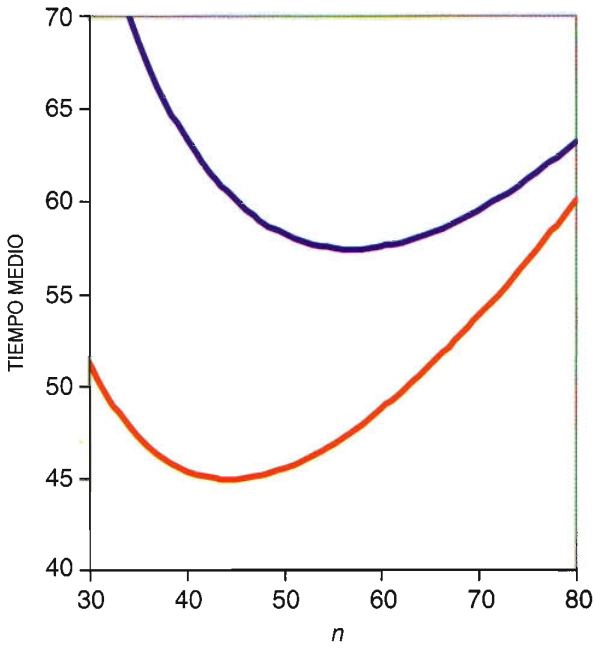
Finalmente, para encontrar el tiempo total medio, tenemos que promediar $Tv + k$ utilizando las probabilidades anteriores para k y v. El cálculo no es sencillo y requiere conocer algunas fórmulas sobre la suma de progresiones geométricas. Sin embargo, el resultado no es excesivamente complicado:

$$t = \frac{Tp^n + n}{1 - p^n} - \frac{p}{1 - p}.$$

En la figura 2 se muestra el tiempo en función de n para $T = 50$, es decir, para cuando invertimos un tiempo en dar la vuelta a la manzana igual al tiempo que tardamos en caminar 50 coches. La curva azul corresponde a $p = 0{,}98$, es decir, a una probabilidad del 2 % de que hay un sitio libre. La curva roja corresponde a $p = 0{,}97$. La estrategia óptima viene dada por el mínimo de cada curva. En el caso $p = 0{,}98$, el mínimo se alcanza para $n = 57$. Es decir, tenemos que despreciar todos los lugares libres que están a una distancia del teatro mayor que 57 coches. Actuando de esta forma, damos un promedio de 0,46 vueltas a la manzana, tardamos en aparcar un tiempo medio de 23,11 segundos y aparcamos a una distancia media de 34,35 coches. El tiempo total es la suma de ambas cantidades, que es 57,46 segundos. Es curioso que una pequeña variación en la probabilidad de que un sitio esté libre altere bastante tales cantidades. Para $p = 0{,}97$ (*curva roja*), el mínimo se alcanza en $n = 45$, se dan, en media, 0,34 vueltas antes de aparcar, tardando un tiempo de 17,02 segundos, y se aparca a una distancia media de 27,98 coches. El tiempo total es entonces 45 segundos.

Es de esperar que, si la probabilidad de encontrar un sitio libre aumenta, entonces el n a partir del cual tenemos que empezar a buscar aparcamiento deberá disminuir. Y así ocurre, porque el mínimo de las curvas pasa de ser 57 a 45. También el n óptimo crece al hacerlo T, ya que en este caso llegar hasta el final de la calle sin haber encontrado aparcamiento tiene una penalización cada vez mayor. La forma concreta en que el n óptimo depende de T y p es bastante complicada, pero, con un poco de manipulación matemática, se puede demostrar que n crece como el logaritmo de T, para una probabilidad p fija.

Como hemos visto, incluso el modelo más simple del problema del aparcamiento presenta complicaciones matemáticas. El problema del aparcamiento recuerda a otro problema clásico: aquel en que alguien debe decidirse entre quedarse con algo o arriesgarse a que aparezca algo mejor. Se trata del problema de *la dote del sultán*. Un sultán decide conceder la mano de una de sus cien hijas a uno de sus súbditos, pero a condición de que supere la siguiente prueba. Las hijas irán pasando, una a una, ante el pretendiente y se



2. Tiempo total medio que tardamos en aparcar y llegar andando al teatro, en función del número n de coches a partir del que comenzamos a buscar aparcamiento. En ambos casos se tarda un tiempo $T = 50$ en dar una vuelta a la manzana. La curva azul corresponde al caso en que la probabilidad de que un sitio esté ocupado es $p = 0{,}98$ y la roja a $p = 0{,}97$.

anunciará la dote que acompaña a cada una de ellas. En cada una de estas ocasiones, el pretendiente tendrá que decidir si se queda con la hija que tiene ante sí o continúa el juego. El pretendiente sólo tendrá el beneplácito para la boda si elige a la hija con la mayor dote.

Las posibles estrategias del pretendiente son muy parecidas a las del buscador de aparcamiento. Tendrá que dejar pasar un cierto número n de hijas y después quedarse con la primera cuya dote sea mayor que la de cualquiera de las anteriores. Si n es pequeño, lo más probable es que se precipite en la decisión pero, si n es muy grande, es muy posible que el pretendiente desprecie a la de mayor dote por salir entre las n primeras. Existe pues un n óptimo que maximiza la probabilidad de elegir a la hija de mayor dote. Se puede demostrar que, para cien hijas, el n óptimo es 37. Es decir, el pretendiente tiene que esperar hasta haber visto a 37 hijas y después elegir a la primera cuya dote sea mayor que las de todas las precedentes (incluidas, por supuesto, las 37 primeras). Con esta estrategia, la probabilidad de acertar es bastante alta, del 37,1 %. Los lectores pueden tratar de resolver este problema. Aunque les advierto de que la solución no es sencilla e involucra algunos conceptos matemáticos no muy conocidos, como los llamados *números armónicos*. Por ello, el próximo mes lo analizaremos con algún detalle.

<parr@seneca.fis.ucm.es>

Juan M.R. Parrondo

La dote del sultán

El mes pasado terminábamos la sección proponiendo a los lectores un problema ya clásico en matemáticas: *la dote del sultán*. Un sultán tiene cien hijas y decide dar la mano de una de ellas a uno de sus súbditos si supera la siguiente prueba: Las hijas desfilarán ante el pretendiente, al que se le anunciará la dote que acompaña a cada una de ellas. El súbdito sólo podrá casarse con la hija de mayor dote si adivina cuál de ellas es. Para ello debe decidir ante cada una de las hijas si la elige para esposa o si prefiere continuar viendo al resto. Una vez rechazada una de las hijas, la decisión no se puede cambiar. Supondremos que todas las dotes son distintas y que el pretendiente no tiene ninguna información previa acerca de su cuantía. ¿Cuál es entonces la mejor estrategia para superar la prueba?

Una estrategia bastante general consiste en dejar pasar un cierto número n de hijas y después elegir la primera cuya dote supere a todas las precedentes (incluyendo las n primeras, obviamente). El problema consiste entonces en calcular el número n que maximiza la probabilidad de elegir la dote más alta.

Analicemos el problema con sólo 4 hijas. Si $n = 0$, elegiremos siempre a la primera de las hijas. Si $n = 1$, no elegiremos en ningún caso a la primera y luego esperaremos hasta que aparezca una con una dote mayor que ella. Las otras dos estrategias posibles son similares, pero con $n = 2$ o $n = 3$. Esta última, $n = 3$, es en realidad muy sencilla: puesto que se rechazan las tres primeras hijas, con esta estrategia elegiremos siempre a la última de ellas. En la tabla se analizan todas las posibilidades.

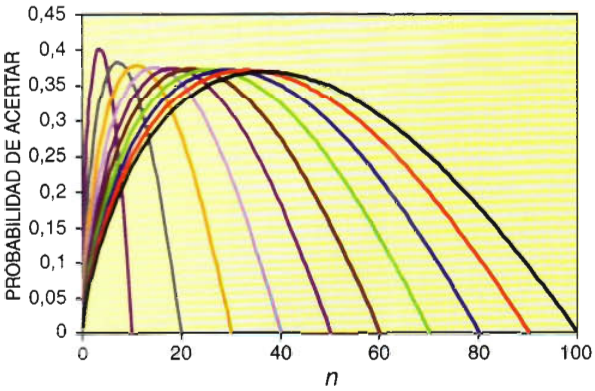
Las cuatro hijas se numeran de mayor (1) a menor dote (4). En la columna de la izquierda se presentan todas las posibles ordenaciones en las que pueden aparecer las cuatro hijas, resaltando en naranja a la de mayor dote. En el resto de las columnas se indica la hija elegida al utilizar cada una de las cuatro posibles estrategias y se destacan en verde los aciertos. Si, por ejemplo, las cuatro hijas aparecen en el orden 4, 1, 3, 2, la estrategia con $n = 0$ elegirá a la primera de ellas, es decir, a la 4, que es la de menor dote. La estrategia con $n = 1$, rechaza a la 4 y, al aparecer la 1 en segundo lugar, la acepta, superando la prueba. Sin embargo, las estrategias con $n = 2$ o $n = 3$, rechazan la 1 y se quedan ambas con la hija número 2.

Para el caso de 4 hijas, la estrategia óptima es la $n = 1$, con la que se alcanza una probabilidad de ganar 11/24 = 45,83 % (24 es el número total de casos u ordenaciones). Es una probabilidad bastante alta si la comparamos con la probabilidad de ganar si se elige una de las hijas al azar, que sería 1/4 = 25 %, la misma que tienen las estrategias con $n = 0$ y $n = 3$. Como vemos, la estrategia óptima se alcanza para un número n medio. Si n es muy pequeño, nos precipitamos y elegimos demasiado pronto, cuando es aún poco probable que haya aparecido la hija de mayor dote. Por otro lado, si n es demasiado alto, lo más probable es que la hija de mayor dote aparezca entre las n primeras y la rechacemos.

Veamos cuál es la probabilidad de acertar con N hijas habiendo rechazado a las primeras n. Acertamos si la de mayor dote está en el puesto $n + 1$ y esto ocurre con probabilidad $1/N$. También acertamos si la mayor está en el puesto $n + k + 1$, lo que ocurre con probabilidad $1/N$, y la mayor de las $n + k$ precedentes está entre las n primeras, lo que ocurre con probabilidad $n/(n + k)$. Por tanto, esta posibilidad ocurre con una probabilidad

$$\frac{1}{N}\,\frac{n}{n+k}\,.$$

Combinación				n=0	n=1	n=2	n=3
1	2	3	4	1	4	4	4
1	2	4	3	1	3	3	3
1	3	2	4	1	4	4	4
1	3	4	2	1	2	2	2
1	4	2	3	1	3	3	3
1	4	3	2	1	2	2	2
2	1	3	4	2	1	4	4
2	1	4	3	2	1	3	3
2	3	1	4	2	1	1	4
2	3	4	1	2	1	1	1
2	4	1	3	2	1	1	3
2	4	3	1	2	1	1	1
3	1	2	4	3	1	4	4
3	1	4	2	3	1	2	2
3	2	1	4	3	2	1	4
3	2	4	1	3	2	1	1
3	4	1	2	3	1	1	2
3	4	2	1	3	2	2	1
4	1	2	3	4	1	3	3
4	1	3	2	4	1	2	2
4	2	1	3	4	2	1	3
4	2	3	1	4	2	1	1
4	3	1	2	4	3	1	2
4	3	2	1	4	3	2	1
Aciertos				6	11	10	6



1. Probabilidad de acertar en función del número n de hijas rechazadas. Cada curva corresponde a un número total de hijas distinto: N = 10, 20, 30, ..., 100.

La probabilidad de acertar es entonces la suma de estas probabilidades extendida a todos los posibles valores de k, que van de 0 hasta $N - n - 1$, es decir:

$$p_{acertar} = \frac{n}{N}\left(\frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{N-1}\right).$$

Como en el caso de cuatro hijas, tenemos que encontrar el número n que maximiza esta probabilidad. En la figura 1 se puede ver la probabilidad de acertar en función de n para distintos valores de N. Vemos cómo cada una de las curvas tiene un máximo que se alcanza para algún valor de n. Por ejemplo, para $N = 100$ hijas, el valor óptimo es $n = 37$. Es decir, la mejor estrategia consiste en rechazar las primeras 37 hijas y luego aceptar la primera que tenga mayor dote que todas las anteriores. Actuando de esta forma, la probabilidad de ganar la prueba es del 37,10 %.

Para abordar el problema de modo más general, nos pueden ayudar algunos conocimientos avanzados de matemáticas. En teoría de números y en análisis, se han estudiado con bastante detalle los llamados *números armónicos*, que se definen de la siguiente forma:

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}.$$

Es sencillo comprobar que la probabilidad de ganar la prueba del sultán, rechazando las n primeras hijas se puede escribir utilizando los números armónicos:

$$p_{acertar} = \frac{n}{N}(H_{N-1} - H_{n-1}).$$

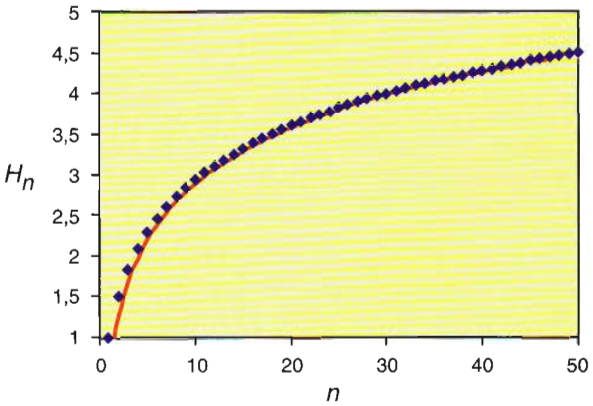
No hay una fórmula sencilla, aparte de la propia definición, que nos proporcione el valor exacto del número armónico H_n. Sin embargo, Euler encontró una fórmula aproximada para dicho valor, que es mejor cuanto mayor sea n:

$$H_n \approx \ln(n) + \gamma.$$

En esta fórmula, ln es el logaritmo neperiano y γ es un número que vale 0,5772 y se llama *constante de Euler-Mascheroni*. Esta constante ejerce un papel importante en distintas ramas y problemas de la matemática, como la teoría de números o el cálculo de series e integrales. La aproximación es bastante buena, como puede apreciarse en la figura 2. Si aplicamos la aproximación de Euler a la probabilidad de acertar, obtenemos:

$$p_{acertar} \approx -\frac{n}{N} \ln\left(\frac{n-1}{N-1}\right).$$

Cuando tanto N como n son muy grandes, la probabilidad de acertar depende sólo de $\alpha = n/N$, es decir, de la fracción de hijas rechazadas de antemano: $p_{acertar} \approx -\alpha \ln \alpha$. A partir de esta expresión, los lectores con algunos conocimientos de análisis matemático podrán encontrar sin mucha dificultad el valor máximo de la probabilidad de acertar: se alcanza para $\alpha = 1/e$ y vale precisamente $p_{acertar,max} = 1/e$. Casualmente, la máxima probabilidad de acertar coincide con el valor óptimo de α. En la solución aparece otro número de



2. Los puntos azules son los números armónicos exactos y la curva roja la aproximación de Euler.

gran importancia en matemáticas y también ligado a Euler: el número e, que vale 2,72828... Su inverso es 1/e = 0,3679..., muy próximo al valor que hemos encontrado para el caso de 100 hijas. En realidad, hemos resuelto el problema suponiendo que α puede tomar cualquier valor, pero en realidad sólo puede tomar valores $\alpha = n/N$, con n entero. La solución final, siempre aproximada, sería el valor posible más próximo a 1/e, que, para $N = 100$, es precisamente $n = 37$.

Es sorprendente la alta probabilidad de acertar con la estrategia óptima. Si elegimos al azar entre 100 hijas, la probabilidad de acertar es del 1 %, mientras que si rechazamos de antemano a las 37 primeras, la probabilidad sube hasta el 37,10 %, que es mayor que un tercio. De modo que el súbdito del sultán puede albergar bastantes esperanzas de obtener la mano de la hija de mayor dote.

<parr@seneca.fis.ucm.es>

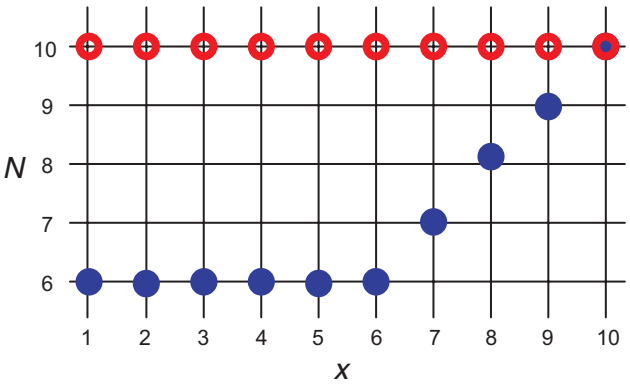
JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Fósiles y lotería

Normalmente jugamos a la lotería comprando un décimo, es decir, tratando de adivinar qué número saldrá de un bombo que contiene un cierto número de bolas. Los paleontólogos se enfrentan en ocasiones a un juego casi opuesto: conocen el número que ha salido de un bombo y tienen que averiguar cuántas bolas hay en él.

Este problema, cuyo planteamiento y parte de su análisis me han sido sugeridos por Antonio Fernández, profesor de matemáticas del Instituto Salvador Allende de Fuenlabrada, parece un tanto trivial. Si, por ejemplo,



1. Diferentes estrategias para adivinar el número *N* de bolas de un bombo del que se ha extraído la bola *x*.

nos dicen que la bola extraída de un bombo es la 8, el número total de bolas *N* puede ser igual o mayor que 8. En principio da la impresión de que no importa por qué número apostemos, con tal de que sea mayor o igual que 8. Sin embargo, esto no es siempre cierto.

Veamos un caso muy sencillo en el que supondremos que el número total de bolas *N* se elige al azar entre 6 y 10. Es decir, sólo puede ser 6, 7, 8, 9 o 10, y toma cada uno de estos valores con probabilidad 1/5. En el bombo se introducen *N* bolas numeradas de 1 a *N* y se extrae una de ellas. La probabilidad de que el número de bolas sea *N* y la bola extraída sea *x* es entonces:

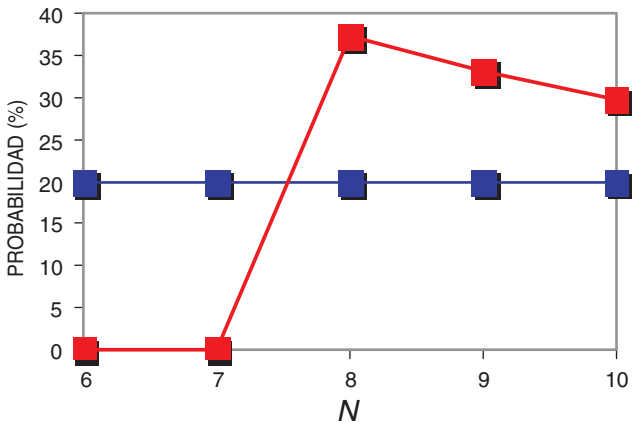
$$P(N,x) = \frac{1}{5} \times \frac{1}{N}$$

siempre que $x \le N$ y $6 \le N \le 10$. Analizaremos primero algunas estrategias sencillas para comprobar que no todas tienen la misma probabilidad de ganar. La primera es apostar siempre por $N = 10$, independientemente del número extraído. La probabilidad de ganar es entonces $P_{\text{ganar}} = 1/5 = 20\,\%$, puesto que no utilizamos la información de la bola extraída *x* y *N* será igual a 10 una de cada cinco veces. Esta estrategia se puede representar en una cuadrícula en donde están todos los posibles pares (x,N), como en la figura 1. Los círculos rojos representan la estrategia de apostar siempre por $N = 10$, independientemente del valor de *x*. La probabilidad de ganar, $P_{\text{ganar}} = 20\,\%$, puede también obtenerse sumando lo que vale $P(N,x)$ en los 10 círculos rojos de la figura 1.

Pero es posible imaginar otras estrategias. Por ejemplo, en la representada por los círculos azules de la figura 1, se apuesta por el *N* más pequeño compatible con la bola extraída *x*, es decir, $N = 6$ si $x = 1, 2, \ldots, 6$, y $N = x$ si $x = 7, 8, 9$ o 10. La probabilidad de ganar con esta estrategia se obtiene sumando las 10 probabilidades $P(N,x)$ de los puntos azules de la figura, y resulta ser: $P_{\text{ganar}} \approx 29{,}6\,\%$. Esta estrategia es por tanto más eficaz que la anterior. No es difícil demostrar que es la estrategia óptima: Los puntos azules no se pueden mover hacia abajo, puesto que entonces apostaríamos por un número *N* de bolas en el bombo inferior al número extraído *x*, lo cual es absurdo. Pero si movemos cualquiera de los puntos hacia arriba, aumenta *N* y disminuye la probabilidad $P(N,x)$. Por tanto, cualquier desviación de la estrategia de los puntos azules tiene como consecuencia una disminución en la probabilidad de ganar.

Es curioso que la mejor estrategia consista en decir que el bombo tiene el menor número de bolas compatible con el número extraído *x*. Mucha gente piensa que la bola extraída *x* es una estimación razonable de la *mitad* del tamaño del bombo y que se debería apostar por algún número *N* cercano a $2x$ y compatible con las reglas del juego. Sin embargo, acabamos de demostrar que esta idea no es correcta.

Un método muy general de resolver problemas como éste es la llamada *inferencia bayesiana*. Este método



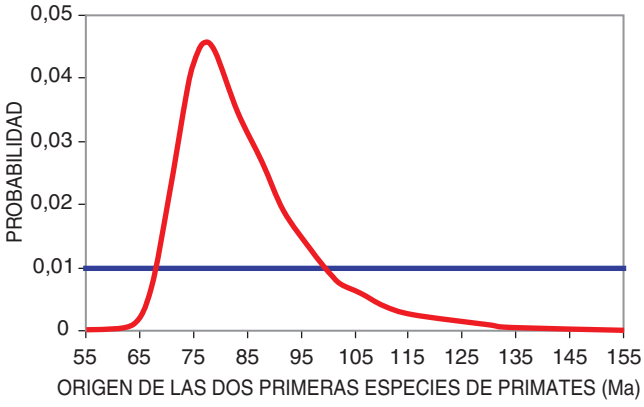
2. Probabilidad *a priori* (*en azul*) y *a posteriori* (*en rojo*) en el problema del bombo, supuesto que se ha extraído la bola $x = 8$.

parte de una cierta distribución de probabilidad para una cantidad desconocida, que se llama *probabilidad a priori.* En nuestro caso la distribución de probabilidad *a priori* de N es $P(N) = 1/5$ para $N = 6,7, 8, 9$ o 10. Cuando se nos muestra el valor de x, la probabilidad de N ya no es la misma. ¿Cuál es esta nueva probabilidad o probabilidad *a posteriori* $P(N|x)$? Su cálculo no es muy difícil. El punto de partida es la probabilidad $P(N,x)$ de que el número total de bolas sea N y que el número extraído sea x. Si, por ejemplo, el número extraído es $x = 8$, N sólo puede tomar los valores 8, 9 y 10 y la probabilidad con la que los toma, es decir, la probabilidad *a posteriori*, tiene que ser proporcional a $P(N,8)$. La probabilidad *a posteriori* para $N = 8$, 9 y 10, supuesto que el número extraído es $x = 8$, es entonces:

$$P(N|x = 8) = \frac{P(N,8)}{P(8,8) + P(9,8) + P(10,8)}$$

En la figura 2 podemos comparar las probabilidades *a priori* y *a posteriori* para el caso $x = 8$. Como vemos, la probabilidad *a posteriori* está más concentrada y alcanza su máximo para $N = 8$, que es precisamente la estrategia de los puntos azules de la figura 1.

Pero, ¿qué tiene que ver este problema con la paleontología? La respuesta guarda relación con una cuestión que me ha sugerido Marián Beltrán, profesora de antropología de la Universidad de Castilla La Mancha. Supongamos que encontramos el fósil de una especie. Si después de datar el fósil, se encuentra que el animal del que proviene vivió hace 8 millones de años, ¿cuándo se originó dicha especie? Si la especie se originó hace N millones de años y si datamos con una precisión de un millón de años, encontrar el fósil es exactamente lo mismo que extraer un número de un bombo con N bolas. En nuestro ejemplo, la bola extraída es $x = 8$ millones de años. Como hemos visto, y aunque sea en cierto modo contrario a la intuición, lo más probable es que la especie se haya originado hace precisamente 8 millones de años. Aunque este "más probable" hay que tomarlo con bastante precaución: como se ve en la figura 2, las probabilidades *a posteriori* para el número N decrecen muy lentamente.



3. Probabilidad *a priori* (*en azul*) y *a posteriori* (*en rojo*) del momento en el que se originaron las primeras especies de primates.

Esta es por supuesto una simplificación muy drástica de los problemas a los que se enfrenta la paleontología, ya que normalmente se dispone de varios fósiles de la misma especie y de otros datos, como el parentesco con otras especies conocidas. Pero la idea básica es similar a la inferencia bayesiana en el problema del bombo. En abril de 2002, el matemático de la Universidad del Sur de California Simón Tavaré y sus colaboradores publicaron en *Nature* un estudio de los fósiles de las distintas especies de primates utilizando técnicas similares a la inferencia bayesiana. La siguiente tabla muestra el número de especies de primates de las que se han encontrado fósiles en función de su antigüedad (en millones de años):

EPOCAS	TIEMPO (en Ma)	NUMERO DE ESPECIES
PLEISTOCENO	0,15	19
	0,9	28
	1,8	22
PLIOCENO	3,6	47
	5,3	11
MIOCENO	11,2	38
	16,4	46
	23,8	36
OLIGOCENO	28,5	4
	33,7	20
EOCENO	37	32
	49	103
	54,8	68

Como se ve en la tabla, el fósil más antiguo encontrado tiene casi 55 millones de años de antigüedad. Sin embargo, estudios de biología molecular indican que los primates provienen de dos especies que se originaron hace unos 90 millones de años. En la figura 3 (tomada de un artículo de Plagnol y Tavaré), podemos ver la probabilidad *a priori* y *a posteriori* del tiempo en el que aparecieron las primeras especies de primates. La probabilidad *a priori* es simplemente una probabilidad uniforme entre 55 y 155 millones de años (en este *a priori* ya se ha tenido en cuenta que el fósil más antiguo tiene 54,8 millones de años de antigüedad), mientras que la probabilidad *a posteriori*, calculada utilizando todos los datos de la tabla y alguna suposición adicional, alcanza su máximo en torno a los 80 millones de años y da una alta probabilidad a una antigüedad de 90 millones de años. Por lo tanto, el estudio de Tavaré serviría para reconciliar el registro fósil con los estudios de biología molecular y avanzar un paso más en el conocimiento de nuestro propio pasado.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

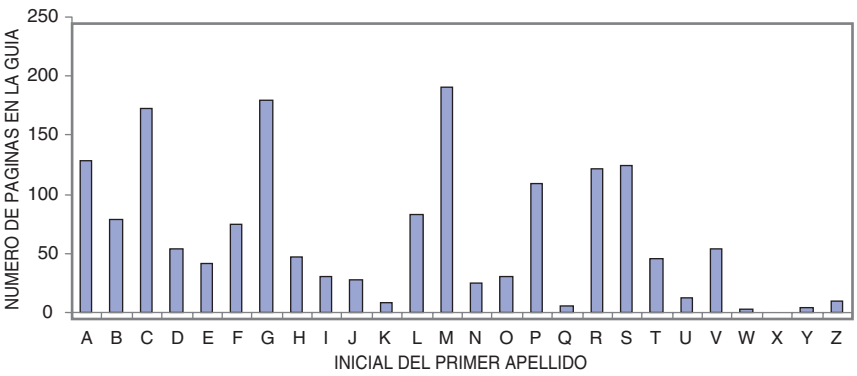
Sorteos polémicos

En ocasiones, las administraciones públicas se ven obligadas a diseñar sorteos con importantes repercusiones para los ciudadanos. Algunos de ellos han sido polémicos por un uso incorrecto de las leyes de la probabilidad. Hace unos años tuvo cierta repercusión un sorteo realizado por el Ministerio de Defensa para determinar quiénes se librarían del servicio militar: los llamados "excedentes de cupo". En el sorteo se asignaba un número de forma aleatoria a cada individuo y se extraía luego en un acto público un número de unos bombos. A partir de dicho número se empezaban a contar los excedentes. El sistema de bombos resultó un fiasco porque los números mayores tenían una probabilidad más alta de salir, debido a un error bastante evidente en el método de extracción. El caso ocupó las páginas de los periódicos e incluso hubo interpelaciones parlamentarias pidiendo al gobierno la repetición del sorteo. No hubo necesidad de ello, puesto que la asignación inicial de números había sido aleatoria (aunque no pública), con lo que el proceso completo era equitativo. No vamos a entrar en detalles sobre aquel pequeño escándalo probabilístico (los lectores interesados pueden encontrar en el número de febrero de 1998 de la revista *Suma* un estudio muy completo del caso, realizado por Roberto Marcellán), sino que analizaremos otro sorteo polémico más actual y que afecta enormemente a muchas familias. Se trata de los métodos que las comunidades autónomas utilizan para distribuir las plazas escolares de los colegios públicos. Ramón Muñoz Tapia, profesor de física de la Universidad Autónoma de Barcelona, me ha enviado un análisis de sus deficiencias que deja en evidencia los escasos conocimientos de teoría de probabilidad que exhibe la administración.

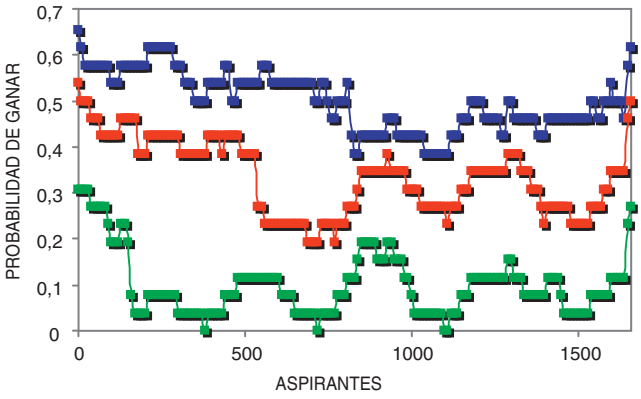
En muchos colegios públicos hay más solicitudes que plazas libres. Para conceder dichas plazas, existe un baremo que tiene en cuenta la proximidad geográfica, el número de hermanos en el centro y otros criterios. Aun así, se dan numerosos empates, porque los criterios del baremo son escasos y la mayoría de los aspirantes cumplen algunos de ellos. Para dilucidar estos empates el gobierno de la comunidad autónoma elige una letra en un sorteo público. Se comienza entonces a asignar plaza a los aspirantes cuyo primer apellido comienza con dicha letra y se continúa por orden alfabético. Si en la asignación se llega al final de la lista alfabética de aspirantes, se continúa el proceso por el comienzo de dicha lista. Finalmente, cuando en un colegio se cubren todas las plazas libres, los siguientes aspirantes son asignados al centro que eligieron como segunda opción en su solicitud, y así sucesivamente.

Este tipo de sorteo es claramente injusto, a pesar de que se utiliza en muchas comunidades autónomas. Alguien llamado Alejandro González Zunzunegui, por ejemplo, elegirá después de todos los García y todos los González independientemente de la letra que salga en el sorteo. Como estos apellidos son muy comunes, es posible que Alejandro se quede sin plaza en el colegio que haya elegido, sea cual sea el resultado del sorteo. Se puede objetar que el número de plazas libres frente al número de empates no es tan pequeño como para que se agoten con una sola letra. Normalmente el número de plazas libres está entre 1/3 y 1/2 del número de empates. En cada colegio se admitiría por tanto más o menos a la mitad de los aspirantes empatados; se recorrería la mitad de la lista alfabética. Si es así, las irregularidades en la distribución de apellidos apenas deberían afectar al reparto. Sin embargo, un análisis más minucioso indica que este argumento no es correcto, incluso con fracciones de aceptación de 1/2 y 1/3.

Para realizar este análisis es necesario conocer la distribución de apellidos según su inicial. Una estimación razonable se puede obtener de la guía telefónica. En la figura 1 se muestra el número de páginas que la guía telefónica de Madrid dedica a cada letra del alfabeto (he eliminado la "ñ", por la que empiezan muy pocos apellidos). Como la guía tiene 1660 páginas, supondré una población de 1660 ciudadanos que han de someterse al sorteo de la primera letra, con apellidos distribuidos igual que en la guía. Es decir, habrá tantos individuos empezando por una letra dada como páginas de la guía correspondientes a dicha letra. Finalmente, para realizar el sorteo hay que fijar la fracción p de plazas en litigio con respecto al número de aspirantes en cada centro, fracción que supondré igual para todos los centros. Con todas estas hipótesis, que no se alejan excesivamente de la realidad, se puede



1. Número de páginas en la guía telefónica correspondientes a las distintas letras del alfabeto.



2. Probabilidad de obtener plaza en el centro elegido en primera opción para los aspirantes ordenados por orden alfabético y para distintos valores de p: 1/2 (*puntos azules*), 1/3 (*puntos rojos*) y 1/10 (*puntos verdes*).

calcular la probabilidad de que un aspirante "gane" el sorteo, es decir, que obtenga plaza en el colegio elegido como primera opción en su solicitud. Para ello basta identificar qué aspirantes ganan con cada letra, algo que depende de la fracción p. Por ejemplo, si $p = 1/2$ y en el sorteo sale la letra A, ganarán los primeros 830 aspirantes en orden alfabético. Si sale la letra L, ganarán los últimos de la lista a partir del 842 (que es el primer aspirante que empieza por L) y los 12 primeros de la lista, porque, al llegar al final de lista sin haber repartido todas las plazas se continúa por el principio de la misma. Finalmente, para calcular la probabilidad de ganar de un determinado aspirante, se cuenta el número de letras con las que dicho aspirante gana y se divide por 26 (que es el número total de letras). El cálculo es un poco tedioso, pero se puede realizar con algún programa informático.

El resultado se muestra en la figura 2, para distintos valores de p: 1/2 (*puntos azules*), 1/3 (*puntos rojos*) y 1/10 (*puntos verdes*). La figura muestra que el sorteo dista bastante de ser equitativo. Para $p = 1/2$, todos los aspirantes deberían tener una probabilidad 1/2 de ganar. Sin embargo, los diez primeros aspirantes (ordenados alfabéticamente) gozan de una probabilidad del 65 %, mientras que los que se encuentran entre los puestos 830 y 850, es decir, con apellidos que empiezan por K y L, o entre 1040 y 1110, franja que corresponde a los últimos puestos de la M y los primeros de la N, tienen una probabilidad de ganar del 38,5 %, porque sólo ganan con 10 letras. Aunque la curva es bastante irregular, se puede explicar cualitativamente mediante los datos representados en la figura 1. Los diez primeros aspirantes de la lista ganan con 17 letras: la A y de la K en adelante. La razón es que la mayoría de los apellidos se concentran en la primera mitad del abecedario. Por ese mismo motivo, los aspirantes que están en la mitad de la lista son los más perjudicados en el sorteo.

Si la fracción p de plazas a repartir es menor, por ejemplo $p = 1/10$, las diferencias se acentúan. Los más beneficiados siguen siendo los primeros en lista, porque las letras que preceden a la A, es decir, las últimas letras del abecedario (recordemos que cuando se llega a la Z en el proceso de reparto se vuelve a comenzar por el principio de la lista), son bastante raras. En esta ocasión hay incluso aspirantes que no pueden ganar nunca: son los últimos con apellidos que empiezan por C, G y M. Ello se debe a que el número de aspirantes que empiezan por estas tres letras es 172, 179 y 191, respectivamente, que son cifras superiores al número total de plazas a repartir: 166.

El sorteo de la letra es, por tanto, manifiestamente injusto. En algunas comunidades autónomas hubo quejas y se mejoró el sistema sorteando las dos primeras letras del apellido en lugar de sólo una. Esta modificación, sin embargo, no es suficiente para que el sistema sea completamente equitativo. Este año, la Generalidad ha dispuesto un sorteo aparentemente justo: se numeran todos los solicitantes de la comunidad autónoma por orden alfabético y, en sorteo público, se obtiene un número correspondiente a uno de ellos, a partir del cual se comienzan a aceptar solicitudes en cada centro. Con este sistema se eliminan las injusticias que hemos analizado aquí.

Sin embargo, hay otra nada desdeñable. El método sería justo si las distribuciones de apellidos en los aspirantes de cada colegio fueran iguales a la distribución en la totalidad de aspirantes de Cataluña. Imaginen que en un determinado colegio todos los aspirantes se llaman García. En ese caso, los últimos de la lista de aspirantes a plaza en ese colegio tendrían una probabilidad aproximadamente $1/S$ de obtenerla, siendo S el número total de aspirantes en toda Cataluña, mientras que los primeros la obtendrían con probabilidad $1–1/S$. El número S suele ser muy alto. Este año está en torno a 100.000. Por eso la ventaja de los primeros García frente a los últimos en nuestro singular colegio sería de 0,99999 a 0,00001. ¡Un sesgo considerable! Aunque éste es un ejemplo extremo, cualquier diferencia entre la distribución de apellidos en un colegio y en toda la comunidad autónoma producirá sesgos en el reparto. Estas diferencias pueden ser bastante apreciables, ya que hay barrios con una mayor concentración de población extranjera o con abundancia de ciertos apellidos. Consultando Idescat, el servicio estadístico de la Generalidad, se puede comprobar que tales diferencias existen. El apellido García, por ejemplo, lo tienen un 24,29 % de residentes en Cataluña, pero en una comarca el porcentaje se eleva hasta casi un 31 % mientras que en otra desciende hasta el 7,15 %. Otros apellidos presentan valores más extremos, como Masip, que llega al 20,91 % en el Priorato mientras que sólo alcanza un 0,39 % en toda Cataluña. Es de esperar que estas diferencias sean mucho más acusadas en poblaciones pequeñas, como la correspondiente a los aspirantes que viven en la misma zona de un municipio.

En definitiva, parece inevitable eliminar los sesgos en cualquier método de reparto que se base en una ordenación alfabética. Lo más justo sería ordenar al azar a todos los aspirantes y extraer de los bombos el número a partir del cual se empieza a elegir plaza. Teniendo informatizadas todas las solicitudes, este sistema no sería difícil de llevar a la práctica. Eso sí, requeriría que nuestros gestores tuvieran algunos conocimientos de probabilidad, o que se dejaran asesorar por mejores expertos en la materia.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La forma de un iceberg

Todos sabemos que la mayor parte de la masa de un iceberg se encuentra sumergida. La "punta del iceberg", es decir, su parte visible, es algo menos de un 10 % del volumen total.

Hace unos meses, paseando por un río parcialmente helado, comencé a lanzar bloques de hielo al agua. Pronto me di cuenta de que ninguno de ellos se parecía a la imagen de la figura 1. Intenté esculpir el hielo para que adquiriera la forma del iceberg prototípico pero, al introducirlo en el agua, el bloque se orientaba siempre con su lado más ancho hacia la superficie. En otras palabras, un trozo de hielo como el de la figura 1 giraría unos 90 grados hasta que el lado más largo quedara horizontal. Según mis "experimentos" la imagen clásica del iceberg es imposible. ¿Es esto cierto?

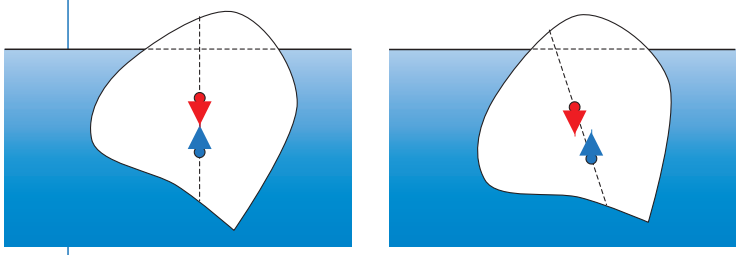
1. Creación artística de un iceberg.

Todos nos imaginamos a Arquímedes envuelto en una toalla y gritando Eureka por las calles de Siracusa. Sabemos que un cuerpo sumergido en agua experimenta una fuerza o empuje hacia arriba igual al peso del agua que desaloja. Si un cuerpo de volumen V tiene una densidad d inferior a la del agua (supondremos que la densidad del agua es igual a 1 gr/cm^3 y mediremos la densidad d del cuerpo en estas unidades), el cuerpo flotará. Cuando el cuerpo llega a la superficie, se alcanza una situación de equilibrio en la que parte de su volumen sale al exterior y queda sumergido un volumen $V_{sum} < V$. En la situación de equilibrio, el peso del cuerpo V_d tiene que ser igual al empuje V_{sum} y, por tanto, la fracción de volumen sumergido V_{sum}/V coincide con la densidad d del cuerpo expresada en gramos por centímetro cúbico. En el caso del hielo, la densidad es aproximadamente 0,9. Por eso la "punta del iceberg" es sólo un 10 % de su volumen total.

La parte difícil del problema es encontrar la orientación que adquiere el cuerpo flotante, una cuestión a la que han dedicado esfuerzos muchos físicos, de Huygens a Sommerfeld. Para resolverla, debemos recordar que el peso de un cuerpo se puede considerar una fuerza aplicada en su centro de masas o centro de gravedad. El peso es en realidad una fuerza que la tierra ejerce sobre todas y cada una de las moléculas del cuerpo, pero se puede demostrar que, a todos los efectos, equivale a una fuerza aplicada en dicho punto. Del mismo modo, el empuje se puede considerar una fuerza aplicada *en el centro de gravedad de la parte sumergida*.

La condición de equilibrio para nuestro cuerpo flotante es sencilla: el centro de gravedad del cuerpo y el centro de gravedad de la parte sumergida tienen que estar en la misma vertical. Si no fuera así, el empuje y el peso harían girar al cuerpo. Esta situación puede verse en la figura 2. En azul se muestra el centro de gravedad de la parte sumergida y el empuje, y en rojo el centro de gravedad del cuerpo y su peso. Si los dos centros de gravedad no están en la misma vertical (*dibujo de la derecha*), el peso y el empuje harán girar al cuerpo.

Sin embargo, en el problema de la orientación del cuerpo hay algo que da una impresión paradójica. Supongamos que giramos ligeramente hacia la izquierda un cuerpo en equilibrio, como en la figura 2. El eje que forman los dos centros de gravedad, el del cuerpo y el de la parte sumergida, habrá girado hacia la izquierda, con lo que el centro de gravedad de la parte sumergida (*punto azul*) quedará a la derecha del centro de gravedad del cuerpo (*punto rojo*). Como se ve en la figura, el peso y el empuje forman ahora un par de fuerzas que hace girar al cuerpo hacia la izquierda, alejándolo de su posición de equilibrio. El argumento puede repetirse para un giro hacia la derecha. Por lo tanto, cualquier pequeña desviación de la orientación de equilibrio crea un par de fuerzas que lo aleja más aún de dicha orientación. Sin embargo, basta observar un cubo de hielo flotando en un vaso de agua para convencerse de que esto no es cierto. Golpee el hielo para que gire un poco y verá que enseguida recupera su situación original. Un cuerpo flotante tiene siempre al menos una orientación de equilibrio estable. ¿Cuál es entonces el error de nuestro argumento? Es bastante simple: hemos supuesto que el centro de gravedad de la parte sumergida gira igual que el cuerpo, algo que no tiene por qué ser cierto, ya que la propia geometría de la parte sumergida cambia. En una orientación de equilibrio estable, el centro de gravedad de la parte sumergida se mueve hacia la izquierda del centro de gravedad del cuerpo cuando éste gira hacia

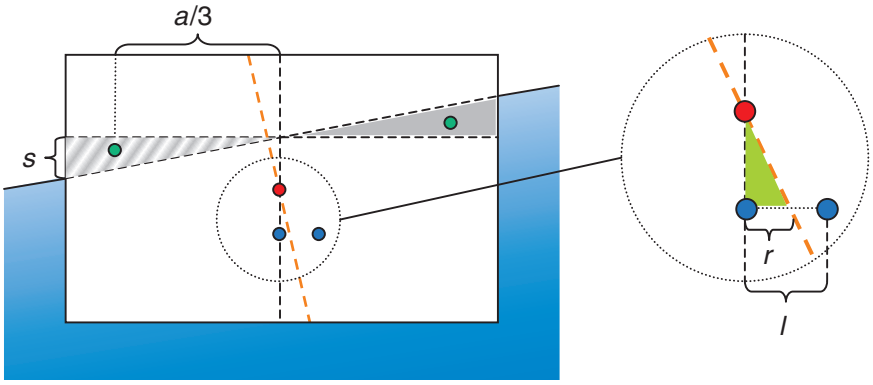


2. La "paradoja" del cuerpo flotante: si giramos el cuerpo hacia la izquierda, aparece un par de fuerzas que hace que el cuerpo gire también hacia la izquierda. ¿Es entonces inestable cualquier orientación del cuerpo?

FOTOMONTAJE DE UWE KILS, http://en.wikipedia.org/wiki/Iceberg (*arriba*)

la izquierda; se crea un par que hace que el cuerpo recupere su orientación de equilibrio. Veamos un caso particular.

Analicemos cómo se puede orientar un rectángulo de lados *a* y *b* y densidad *d*. En la figura 4 se ofrece una posible orientación de equilibrio, en donde los centros de gravedad del cuerpo y de la parte sumergida están alineados verticalmente. Esta no es la única orientación de equilibrio: girando el rectángulo 90 grados, encontraríamos otra orientación posible, cuyo análisis es idéntico, intercambiando los lados *a* y *b*. El centro de gravedad del cuerpo (*punto rojo*) está situado en la mitad del rectángulo, es decir, a una altura *b*/2 con respecto a su base. La parte sumergida del rectángulo es de nuevo un rectángulo de altura *bd*, puesto que, así, la fracción de volumen sumergido es igual a la densidad del cuerpo *d*. Por lo tanto, el centro de gravedad de la parte sumergida (*punto azul*) está a una altura *bd*/2. La distancia entre los dos centros de gravedad es entonces $b(1-d)/2$, tal y como se indica en la figura 3.



3. Una posible situación de equilibrio para un rectángulo flotante.

¿Cómo cambian los centros de gravedad si giramos ligeramente el cuerpo hacia la derecha? El giro está representado en la figura 3 desde el punto de vista del cuerpo. Es decir, en la figura he girado la superficie del agua hacia la izquierda. Después del giro, el triángulo rayado, antes sumergido, deja de estarlo; otro triángulo igual, el sombreado en gris, se sumerge, de modo que la fracción de volumen sumergida no cambia, pero sí su forma. En consecuencia, el centro de gravedad de la parte sumergida se desplaza hacia la derecha una cantidad *l*, como se indica en la ampliación del dibujo. El nuevo centro de gravedad se puede calcular como el promedio, convenientemente ponderado, de tres partes: la parte que estaba sumergida al principio, el triángulo rayado (con peso negativo) y el triángulo gris. Con ello se obtiene:

$$l = 2 \times \frac{a/3 \times A_\Delta}{abd} = \frac{as}{6bd};$$

$A_\Delta = as/4$ es el área del triángulo rayado; hemos utilizado el hecho de que los centros de gravedad de los triángulos (*puntos verdes en la figura 3*) están a una distancia *a*/3 del eje vertical. Para que la orientación inicial sea de equilibrio estable, se requiere que el par de fuerzas que aparece después del giro sea hacia la izquierda, de modo que tienda a restituir la orientación inicial. Eso ocurrirá si el nuevo centro de gravedad de la parte sumergida queda a la derecha del centro de gravedad del cuerpo, es decir, si $l > r$, ya que línea naranja de la figura 3 constituye la nueva vertical. La condición de estabilidad es entonces:

$$\frac{as}{6bd} > r \Rightarrow \frac{a}{6bd} > \frac{r}{s}.$$

Finalmente, el cociente *r*/*s* puede expresarse en términos de los datos del problema si se tiene en cuenta que el triángulo rayado y el verde son equivalentes. Así que:

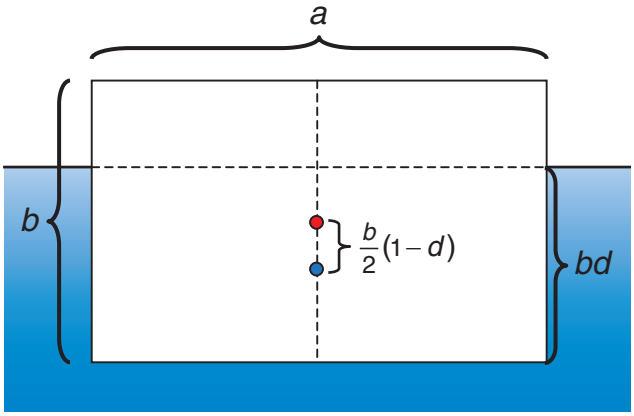
$$\frac{r}{s} = \frac{b(1-d)/2}{a/2} = \frac{b(1-d)}{a}$$

y la condición de equilibrio es:

$$\left(\frac{a}{b}\right)^2 > 6d(1-d).$$

Para el hielo ($d = 0{,}9$) esta condición es $a/b > 0{,}73$. Vale decir: cualquier orientación con $a > b$ es estable. Pero un rectángulo en el que la relación entre el lado menor y el mayor sea superior a 0,73, se puede orientar de forma estable tanto con su lado mayor horizontal como vertical. Dependiendo del valor de *d*, pueden aparecer otras orientaciones estables, e incluso asimétricas, con sólo una o tres de las esquinas del triángulo sumergidas.

Sin embargo, la anchura del bloque de hielo de la figura 1 es aproximadamente la mitad de su altura. Por lo tanto, la orientación del fotomontaje no es estable, al menos si el iceberg es homogéneo. Debo indicar que el problema de la orientación de un iceberg reviste mayor complejidad que la abordada aqui. Depende incluso de su historia geológica. Por ejemplo, se sabe que la parte sumergida se funde más rápido que la visible; por ello, el iceberg puede volcar de modo repentino. No obstante, creo que nuestro análisis ofrece razones suficientes para dudar de la imagen que la mayoría nos hacemos de estos grandes bloques de hielo.



4. Cómo cambia el centro de gravedad de la parte sumergida cuando se gira ligeramente el cuerpo hacia la izquierda.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

¿Amigos para siempre?

Desde que John von Neumann y Oskar Morgensten crearan la teoría de juegos en los años cuarenta, se intenta utilizar las matemáticas para estudiar un asunto tan sinuoso e impreciso como las relaciones humanas. Se crean modelos que simplifican los conflictos reales, pero aún retienen sus aspectos esenciales. Uno de los ejemplos clásicos es el *dilema del prisionero*, que probablemente sea conocido por muchos lectores de *Investigación y Ciencia*.

Recordemos la formulación original del dilema. La policía arresta a dos sospechosos de un robo. No hay pruebas suficientes para condenarlos por ese delito, sino sólo por uno menor, como la posesión ilegal de armas. Ante esta situación, el juez ofrece a cada uno de los dos sospechosos, y por separado, el siguiente trato: "Si tú confiesas y tu cómplice no lo hace, él será condenado a 10 años de cárcel por el robo mientras que a ti te perdonaremos y saldrás libre mañana. Si él confiesa y tú callas, tú pasaras los 10 años en la cárcel y él quedará libre. Si ninguno de los dos confiesa, os encerraremos por posesión de armas 1 año. Finalmente, si ambos confesáis, seréis condenados a 6 años de cárcel". Cada uno de los prisioneros se enfrenta a un peliagudo dilema: "Si callo, puede que me caigan 1 o 10 años, mientras que si confieso puedo salir libre o pasar 6 años en la cárcel. Mi compañero estará pensando lo mismo y lo más probable es que me delate. Por lo tanto, voy a confesar." Razonando de esta forma, ambos prisioneros confiesan y obtienen una pena considerable: 6 años. Si confiaran en la lealtad del compañero, callarían ambos y saldrían bastante bien parados, con sólo 1 año de condena. Pero esta confianza debe ser muy sólida. Si yo supongo que mi compañero me va a ser fiel, ¿por qué no delatarle y salir libre? O peor aún: ¿no caerá él en esta misma tentación y me delatará? Si lo hace pasaré 10 años entre rejas. Luego, lo mejor es prevenir esta situación y confesar. Parece que delatar es la estrategia más segura y, sin embargo, no es la óptima para el conjunto de los dos sospechosos. Si pudieran negociar entre ellos y dar de forma simultánea su respuesta ante el juez, lo más probable es que acordaran callar.

El dilema del prisionero es un ejemplo muy simplificado de una situación común entre personas y organizaciones. Por ejemplo, si falseo mi declaración de la renta para pagar menos impuestos y el resto de los ciudadanos es honrado, está claro que obtengo un cierto beneficio económico, siempre que no me descubran. Pero si todos los contribuyentes defraudasen, los ingresos del estado se reducirían de modo que todos saldríamos perjudicados. Este caso es más complejo que el dilema del prisionero, porque involucra muchas personas en lugar de sólo dos, pero se observa el mismo fenómeno: el defraudador gana solamente si es el *único* insolidario.

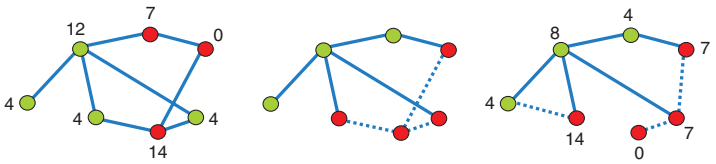
El dilema ha dado lugar a un gran número de trabajos de investigación e incluso se han celebrado torneos en donde compiten programas de ordenador. En el primero de estos torneos resultó ganadora una estrategia de la mayor simplicidad: hacer exactamente lo que ha hecho el contrario en el turno anterior. A pesar de que venciera esta estrategia tan ecuánime, un análisis matemático del dilema del prisionero jugado un gran número de turnos muestra que los jugadores "racionales", es decir, jugadores que tratan de minimizar sus pérdidas, tenderían a defraudar constantemente. En los años noventa, Martin A. Nowak y Robert M. May colocaron a un gran número de estos jugadores racionales en las casillas de un tablero, de modo que cada uno de ellos jugara con sus cuatro vecinos y cambiara de estrategia imitando al vecino más exitoso. Descubrieron que en este caso sí pueden existir de forma estable regiones en donde todos los jugadores cooperan. En 1995, Nowak, May y Sigmund publicaron en nuestra revista un análisis detallado de este modelo.

Más recientemente, los físicos del Instituto Mediterráneo de Estudios Avanzados (IMEDEA) Víctor M. Eguiluz, Martín G. Zimmermann y Maxi San Miguel, junto con el filósofo de la Universidad de las Islas Baleares Camilo J. Cela-Conde, han utilizado el dilema del prisionero para crear un modelo simple de organización social en el que la cooperación es también estable. Para analizarlo, es más conveniente formular el dilema como un juego en el que cada uno de los dos jugadores gana puntos en lugar de años de cárcel. En la tabla siguiente se muestra la puntuación del jugador 1 en cada una de las cuatro posibles situaciones (la puntuación del jugador 2 es la misma sin más que cambiar filas por columnas):

		ESTRATEGIA DEL JUGADOR 2	
		COOPERAR	DEFRAUDAR
ESTRATEGIA DEL JUGADOR 1	COOPERAR	4	0
	DEFRAUDAR	7	0

En el modelo del grupo del IMEDEA, un gran número de individuos juegan por parejas que se deciden inicialmente al azar. Cada jugador tiene una estrategia: o bien coopera o bien defrauda, que también se elige inicialmente al azar. Como en el sistema de Nowak y May, los jugadores "ven" lo que ha ganado cada uno de sus vecinos y copian la estrategia del que tiene mayor puntuación (los empates se deshacen eligiendo la estrategia al azar). En la figura 1, se puede ver cómo cambian las estrategias en el primer turno (*dibujo del centro*). El lector puede comprobar que, si sólo se aplicaran estas reglas de imitación, la red de la figura 1 acabaría alternando entre dos configuraciones de estrategias en las que conviven jugadores que cooperan y jugadores que defraudan.

Pero el modelo es más interesante si permitimos también a cada jugador escoger "sus amigos", es decir, a aquellos individuos con los que juega. Está claro que



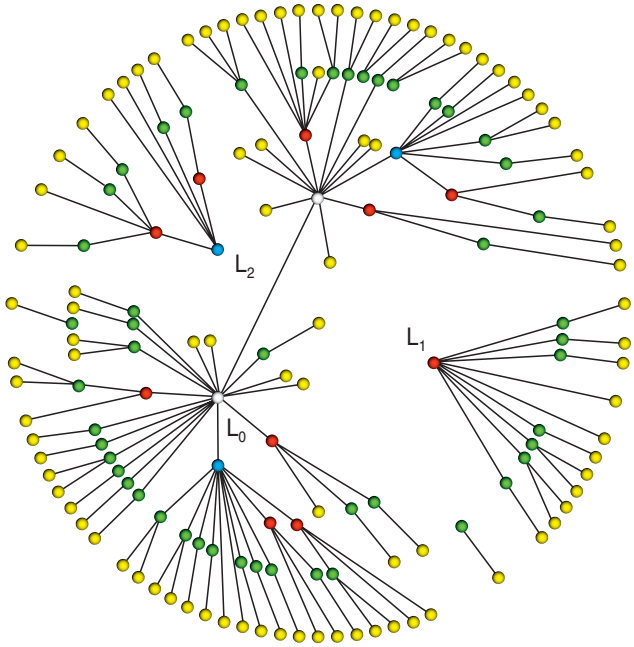
1. Una posible configuración de 6 jugadores que cooperan (*círculos verdes*) o defraudan (*círculos rojos*). En el dibujo de la izquierda se muestra la configuración inicial y las puntuaciones obtenidas en el primer turno. El dibujo central muestra las nuevas estrategias y se han punteado los enlaces "no deseados". En el dibujo de la derecha observamos que dos de estos enlaces han cambiado mientras que uno permanece.

dos jugadores que cooperan estarán satisfechos de estar emparejados. Por otro lado, cualquier jugador querrá "romper relaciones" con otro que defrauda. El modelo admite esta posibilidad de la siguiente forma. Si el más exitoso de los vecinos de un jugador A es un defraudador B, entonces el jugador A rompe su relación con B con una probabilidad p y establece una nueva relación con otro individuo de la sociedad elegido al azar. Observemos que A será siempre un defraudador, porque si no lo era antes de ese turno, en el proceso de copia de estrategias exitosas habrá imitado el comportamiento defraudador de B. Con esta regla la red de conexiones cambia en cada turno, como lo hace en la vida real nuestra red de amistades o colaboraciones, una propiedad de las redes sociales que se denomina *plasticidad social* y que en el modelo puede regularse variando la probabilidad p de ruptura de relaciones.

La plasticidad social diseñada por el grupo del IMEDEA hace que los defraudadores se queden más y más solos, pero a la vez los defraudadores que aparecen por imitación pueden dar con nuevas "víctimas" y aumentar sus ganancias. En la figura 1 podemos ver que el jugador que en el primer turno tenía 14 puntos ha sido imitado por dos cooperadores que han roto sus relaciones con él, bajando hasta 0 su puntuación en el segundo turno. Pero también un jugador que inicialmente cooperaba se ha vuelto defraudador aumentando su puntuación de 4 hasta 14 puntos.

El modelo tiene un comportamiento bastante complicado, pero se pueden deducir algunas de sus propiedades. Por ejemplo, una situación en donde todos cooperan o defraudan es estable porque no hay nadie diferente a quien imitar. No obstante, en el caso en que todos defraudan, las estrategias no cambian, pero sí lo hará constantemente la red, puesto que se estarán rompiendo y creando relaciones sin cesar.

¿Existen configuraciones estables en donde convivan jugadores que cooperan y defraudan? Sí, pero para ello son necesarias ciertas condiciones. La primera es que no puede haber enlaces entre defraudadores, porque acabarían rompiéndose en algún momento. La segunda es que, si un jugador C que coopera está jugando con un defraudador D, entonces la puntuación de C debe ser menor que la de D (si no fuera así D imitaría a C) y, además, C debe tener como vecino a otro jugador L que coopera y que tiene mayor puntuación que D (si no fuera así, C imitaría a D). Es decir, los defraudadores son estables si "explotan" a jugadores que cooperan y



2. Estructura social estacionaria de "cooperantes".

tienen una puntuación baja pero que tienen como vecinos a "cooperantes exitosos".

Las configuraciones estacionarias están entonces formadas por 3 tipos de jugadores: los *líderes*, que son jugadores que cooperan, tienen un gran número de conexiones con otros cooperantes y por ello alcanzan la máxima puntuación entre todos sus vecinos; los *conformistas*, que cooperan, tienen una puntuación menor que alguno de sus vecinos cooperantes, pero no cambian de estrategia; y los *explotadores*, que son defraudadores que juegan con uno o más conformistas que no rompen relaciones con ellos porque tienen también como vecino a un cooperante más exitoso que el propio explotador. En la figura 2 se puede ver la red social que se alcanza cuando se simula en el ordenador el modelo con 10.000 jugadores y una plasticidad $p = 0{,}1$. En la figura sólo se muestran los cooperantes más exitosos. Vemos que aparecen tres líderes, L_0, L_1 y L_2, siendo el primero de ellos el que obtiene mayor puntuación. Según nos alejamos del centro del círculo, los cooperantes tienen menos conexiones y su puntuación es más baja. Como ya hemos dicho, los explotadores, que no se muestran en la figura 2, "cuelgan" de algunos cooperantes que no son líderes y tienen puntuaciones intermedias.

El trabajo del grupo del IMEDEA profundiza mucho más en este modelo. Por ejemplo, muestra cómo la fracción de explotadores aumenta cuando se reduce la plasticidad social. El mismo grupo de investigación ha estudiado la influencia de la plasticidad social en otro modelo que intenta reproducir la difusión de rasgos culturales entre la población, el llamado *modelo de Axelrod*; se demuestra con ello que la plasticidad puede estabilizar un cierto grado de diversidad cultural. Los lectores interesados en este y otros problemas de *matemática social* pueden visitar la página web del grupo: http://www.imedea.uib.es/physdept/eng/lines/social.html

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Cribas y números primos

Uno de los campos más fascinantes de la matemática es el estudio de los números enteros, la llamada *teoría de números*. Dentro de este campo, los números primos ocupan un lugar privilegiado. Su definición es extremadamente simple. Muchos estudiantes de primaria saben qué es un número primo y incluso pueden determinar si un número cualquiera lo es o no. Sin embargo, los números primos guardan secretos que sólo han podido revelarse mediante técnicas matemáticas muy refinadas y aún hoy plantean muchos desafíos. Un ejemplo clásico es la conjetura de Goldbach, que afirma que cualquier número par puede expresarse como la suma de dos primos (por ejemplo, 24 es 19 + 5). A pesar de la simplicidad de su enunciado, nadie ha podido aún demostrarla ni refutarla.

Pero los matemáticos sí han podido adentrarse en algunos de los misterios de los números primos. Por ejemplo, han estimado cuántos números primos hay entre 1 y n, una cantidad que suele denotarse por $\pi(n)$. Ya Euclides demostró que hay infinitos primos, es decir, que $\pi(n)$ crece indefinidamente. Pero, ¿cómo crece? O en otras palabras, ¿cómo están distribuidos los números primos? Por experiencia sabemos que su frecuencia disminuye: hay 168 entre 1 y 1000, 135 entre 1001 y 2000, 127 entre 2001 y 3000, o ya sólo 83 entre 405.001 y 406.000. El descenso es muy lento pero continuado. La *criba de Eratóstenes*, una de las formas clásicas de obtener una lista con todos los números primos, nos permite atisbar las razones de este descenso y de su lentitud.

La criba de Eratóstenes consiste en lo siguiente. Escribimos una lista con todos los números de 1 a N. A continuación tachamos los múltiplos de 2, es decir, los números pares. El primer número no tachado después del 2 es el 3. Tachamos entonces los múltiplos de 3. Buscamos el siguiente número no tachado, que es el 5, tachamos sus múltiplos, y así sucesivamente. De esta *criba* quedan unos pocos números no tachados, que son precisamente todos los primos del 1 al N. El resultado de la criba de los 50 primeros números se puede ver en la siguiente tabla, en donde se muestran en verde los números eliminados en la criba del 2, en naranja los del 3, en azul los del 5 y en rojo los del 7:

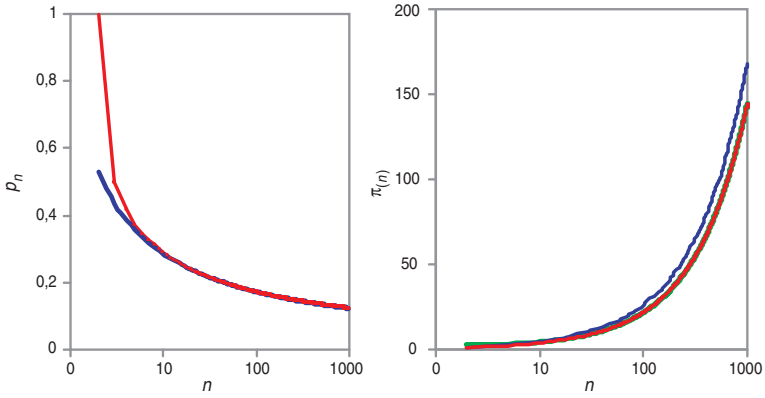
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50

La criba de Eratóstenes nos indica por qué los números primos son más escasos cuanto más avanzamos en la lista. Un número alto puede ser tachado en cualquiera de las cribas de los primos menores que él. Por lo tanto, tiene más "probabilidad" de ser eliminado que un número bajo. Por ejemplo, el 49 se ha salvado de la criba del 2, 3 y 5, pero ha sido eliminado por la criba del 7. Por otra parte, las cribas cada vez eliminan menos números por dos razones: porque tachan menos números y porque las cribas anteriores ya han eliminado muchos de ellos. Es por tanto cada vez menos "probable" que las cribas de números altos eliminen nuevos números. He escrito las palabras "probabilidad" y "probable" entre comillas porque en las cribas de Eratóstenes no hay nada azaroso: se van tachando números en la lista de forma periódica, cada 2, 3, 5,... lugares. Sin embargo, el matemático David Hawkins tuvo la ingeniosa idea de modificar estas cribas de forma que sí fueran aleatorias, dando lugar a los *primos aleatorios de Hawkins*.

La *criba de Hawkins* funciona de la siguiente manera. Como en la de Eratóstenes, se escribe una lista con los números del 1 al N. El primer primo aleatorio es el 2. Lo llamaremos $H1$. A partir del 2 tachamos cada número con una probabilidad 1/2, es decir, lanzamos una moneda en cada número para decidir si lo tachamos o no. Observen que el resultado de esta criba es similar a la de Eratóstenes. Si N es muy grande, habremos tachado aproximadamente la mitad de los números de la lista. Tomamos ahora el primer número mayor que 2 que no haya sido tachado. Será nuestro segundo primo $H2$, que no tiene por qué ser el 3, ya que ha podido quedar eliminado en la criba aleatoria del 2. Supongamos que sí lo es, es decir, que $H2 = 3$. La criba aleatoria correspondiente al 3 consiste en tachar cada número mayor que 3 con una probabilidad 1/3. Como se puede comprobar, en estas cribas aleatorias se tacha, de media, la misma cantidad de números que en las cribas de Eratóstenes, pero, en lugar de hacerlo de forma periódica, se hace de forma aleatoria. La criba continúa de la misma manera. $H3$ es el siguiente número de la lista no tachado. Se toma como tercer primo aleatorio, se tachan los siguientes números de la lista con probabilidad $1/H3$, y así sucesivamente. Los primos de Hawkins son realmente aleatorios, es decir, cambian cada vez que uno repite el proceso completo de criba. En la siguiente tabla pueden ver un ejemplo de criba de Hawkins hecha con los primeros 50 números. El código de colores es similar al de la criba de Eratóstenes: verde para los tachados en la criba de $H1 = 2$, naranja para la criba del segundo primo $H2$, que en este caso ha resultado ser el 3, azul para la de $H3 = 5$ y rojo para la de $H4 = 8$. Hemos necesitado un color más, el amarillo, para la criba del quinto primo $H5 = 9$, que no aparecía en la de Eratóstenes:

1	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
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50

Las tablas resultantes de las dos cribas son muy diferentes, pero comparten algunas características. En la de Eratóstenes hemos obtenido 16 primos, el primero ha eliminado 24 números, el segundo 7, el tercero 2 y el cuarto 1. En la de Hawkins hemos obtenido 14 primos, el primero ha eliminado 25, el segundo 5, el tercero 3, el cuarto 1 y el quinto 2. Los sucesivos primos cada vez eliminan menos números, porque tachan con menor frecuencia o probabilidad y también porque los primos anteriores han eliminado ya bastantes números, como ocurre en la criba de Eratóstenes. La criba de Hawkins reproduce, al menos de forma estadística, los aspectos fundamentales de la criba de Eratóstenes. ¿Podrá entonces ayudarnos a estimar $\pi(n)$? Sí, porque los primos aleatorios se distribuyen de forma parecida a los primos reales y es mucho más fácil calcular cuántos primos aleatorios hay.



En la gráfica de la izquierda se muestra la probabilidad p_n de que n sea primo aleatorio (*en azul*) y se compara con la fórmula aproximada (*en rojo*). En la gráfica de la derecha se muestra $\pi(n)$ (*en azul*), junto con el número medio de primos aleatorios (*en rojo*) y la conjetura de Gauss (*en verde*).

¿Cuál es la probabilidad p_n de que un número cualquiera n sea primo aleatorio? La probabilidad de que n sea primo es muy parecida a la de que lo sea $n + 1$, puesto que ambos han sufrido las mismas cribas antes de llegar a la posible criba de n. Es más, si n resulta no ser primo, la probabilidad de que $n + 1$ lo sea vale precisamente p_n. Por otro lado, si n es primo, $n + 1$ sufrirá su criba y se tachará con una probabilidad $1/n$. En este caso la probabilidad de que $n + 1$ sea primo es p_n, que es la probabilidad de serlo antes de la criba del n, multiplicado por $1 - 1/n$, que es la probabilidad de sobrevivir a la criba del n. Por otra parte, n es primo con probabilidad p_n y no lo es con probabilidad $1 - p_n$. Tenemos entonces la siguiente tabla de probabilidades:

n	$n+1$	Probabilidad
No primo	Primo	$(1 - p_n) \times p_n$
Primo	Primo	$p_n \times p_n\left(1-\frac{1}{n}\right)$

Sumando la última columna, obtenemos la probabilidad de que $n + 1$ sea primo, independientemente de lo que haya sido n:

$$p_{n+1} = p_n\left(1-\frac{p_n}{n}\right).$$

Como $p_2 = 1$, podemos aplicar esta ecuación para obtener las distintas probabilidades p_n. Resulta $p_3 = 1/2$, como cabía esperar, puesto que el 3 sólo sufre la criba del 2; p_4 es 5/12 = 0,42..., p_5 = 215/576 = 0,37..., etc. En la ilustración se pueden ver las probabilidades de ser primo aleatorio para los primeros 1000 números naturales. Encontrar una fórmula general para p_n no es posible; sin embargo, sí puede encontrarse una fórmula aproximada que es mejor cuanto mayor sea n. La fórmula es:

$$p_n \approx \frac{1}{\ln n + 1,2}$$

y se representa en la figura mediante la curva azul, que prácticamente es indistinguible de la roja para números n grandes. Los lectores con conocimientos avanzados de matemáticas pueden deducir esta fórmula aproximando la ley de recurrencia por una ecuación diferencial que se resuelve con facilidad (la constante 1,2 se calcula ajustando la solución a los valores reales de p_n). El número medio de primos aleatorios entre 1 y n será $\pi_{\text{aleat.}}(n) = p_1 + p_2 + ... + p_n$. Si ahora suponemos que los primos reales tienen una distribución parecida a la de los primos aleatorios, encontramos que:

$$\pi(n) \approx p_1 + p_2 + ... + p_n$$

y se puede demostrar que, para n muy grande, esta suma es aproximadamente $n/\ln n$. Este resultado, que $\pi(n)$ sea aproximadamente igual a $n/\ln n$ es uno de los teoremas fundamentales de la teoría de números. Se llama *teorema de los números primos* y fue intuido por Gauss y demostrado más tarde por Riemann. En la figura (*gráfica de la derecha*) se muestra el valor real de $\pi(n)$ junto con la conjetura de Gauss, $n/\ln n$, y $\pi_{\text{aleat.}}(n)$. Comprobamos que las dos últimas curvas prácticamente coinciden y subestiman ligeramente el valor real de $\pi(n)$, aunque esta discrepancia se hace cada vez menor a medida que aumenta n.

La demostración del teorema de los números primos es muy técnica. Sin embargo, la criba de Hawkins ha permitido obtener el mismo resultado de una forma más sencilla, aunque no rigurosa. Más aún, la criba de Hawkins nos proporciona una imagen de por qué los números primos se hacen cada vez más escasos y por qué lo hacen de forma tan lenta. Sanjoy Mahajan, que ha escrito en el Caltech, una muy recomendable tesis doctoral sobre cálculos de orden de magnitud en física en la que se incluye un estudio detallado de los primos aleatorios, considera a éstos un *modelo* de los primos reales, es decir, una versión simplificada de los primos reales que, sin embargo, conserva sus principales características. En esta sección hemos visto en ocasiones modelos matemáticos sencillos de sistemas físicos, económicos o sociológicos. ¡Más sorprendente es que haya *modelos* matemáticos para conceptos también matemáticos!

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Más paradojas de alternancia

Michel Steiver, de la Universidad de Colorado, me ha hecho llegar un ejemplo muy interesante en el que la combinación de dos juegos o inversiones perdedoras da lugar a un juego ganador. Un fenómeno similar a la *Paradoja de Parrondo*, de la que hemos hablado aquí en varias ocasiones. Esta vez el ejemplo está relacionado con situaciones bastante realistas que se dan en mercados financieros como la bolsa o los fondos de inversión.

Imagínense un cierto producto financiero, una acción de una empresa, un fondo de inversión, etc., que cada día da lugar o bien a una ganancia del 30 por ciento o bien a una pérdida del 25 por ciento del capital invertido. Es decir, cada euro invertido se convierte en un día, o bien en 1,30 o bien en 0,75 euros. Supondremos también que cada una de estas dos posibilidades ocurre con la misma probabilidad: 1/2. Este tipo de comportamiento se llama *modelo de árbol binomial* porque cada día pueden ocurrir dos posibilidades que hacen que las posibles evoluciones del precio de la acción a lo largo del tiempo crezcan como las ramas de un árbol. Se trata de un modelo muy simplificado de la evolución del precio de una acción que sirve para enseñar análisis financiero e incluso para hacer algunos cálculos del riesgo de una cierta inversión.

Da la impresión de que comprar esta acción o participar en este fondo es una inversión ganadora, puesto que la ganancia del 30 % es superior a la pérdida del 25 % y ambas posibilidades son igualmente probables. Un análisis más minucioso de cómo evoluciona el capital invertido nos mostrará que la inversión es bastante arriesgada. En la gráfica de la izquierda de la figura vemos seis posibles evoluciones del capital invertido. En cada una de las curvas se ha escogido aleatoriamente cada día la posibilidad ganadora, en la que el capital se multiplica por 1,30, o perdedora, en la que se multiplica por 0,75. Como cada día el capital se multiplica por un factor aleatorio, las fluctuaciones son bastante grandes. Vemos que, partiendo de un euro, una de las curvas sube hasta 10.000 euros. Sin embargo, todas ellas acaban bajando y se puede demostrar que, para un tiempo lo bastante largo, el capital se hace siempre prácticamente nulo.

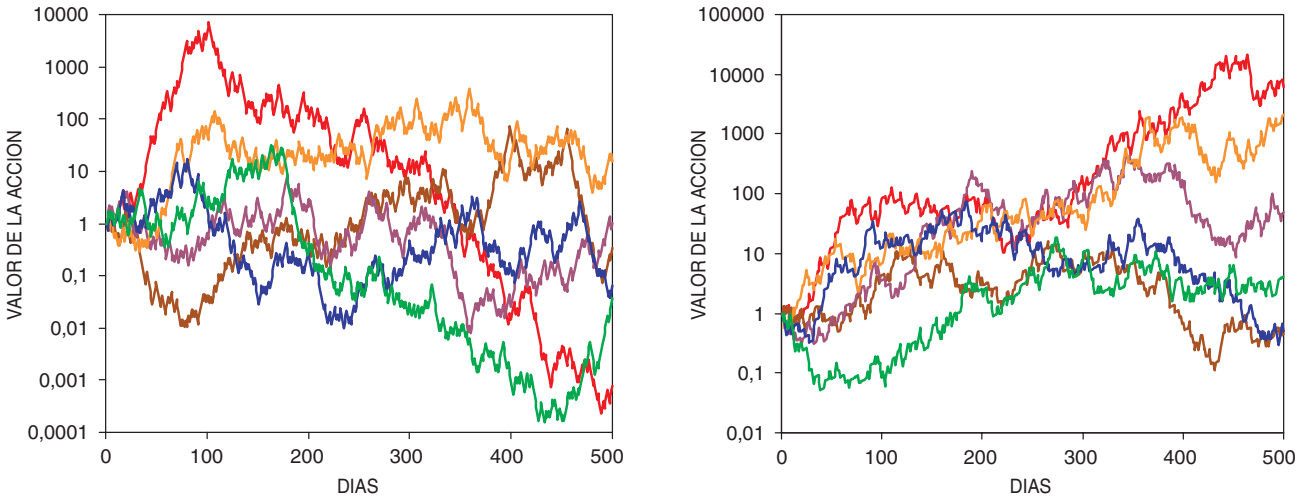
La idea de Steiver es combinar dos inversiones iguales a la anterior. Deben ser además completamente independientes, es decir, en un día una puede subir y la otra bajar, bajar ambas o subir ambas. El precio de cada una de estas acciones sigue una curva idéntica a las de la gráfica de la izquierda de la figura. Sin embargo, y este es el curioso resultado de Steiver, si cada día distribuimos nuestro capital entre las dos acciones por igual, el resultado es el de la gráfica de la derecha en la figura. De nuevo observamos grandes fluctuaciones, pero el comportamiento de esta cartera de inversión combinada es bastante mejor que el del precio de cada acción por separado.

Veamos en detalle qué está ocurriendo. En primer lugar, analizaremos el precio de una sola acción, es decir, el comportamiento de las curvas en la gráfica de la izquierda de la figura; más tarde estudiaremos la inversión combinada. Si inicialmente el precio de la acción es de un euro, tras N días será:

$$P_N = 1{,}3^n \times 0{,}75^{N-n}$$

siendo n el número de días en los que el precio ha subido y $N - n$ el número de días en los que ha bajado. Tras un período largo de tiempo, lo más probable es que la mitad de los días el precio haya aumentado y la otra mitad haya disminuido. En ese caso, $n = N/2$ y el precio es:

$$P_N \sim (1{,}3 \times 0{,}75)^{N/2} = \left(\sqrt{1{,}3 \times 0{,}75}\right)^N \sim 0{,}987^N$$



Capital en función del tiempo para varias realizaciones de las carteras no diversificada (*izquierda*) y diversificada (*derecha*)

en donde hemos aplicado una conocida propiedad de las potencias: elevar a la mitad de un número *N* es lo mismo que extraer la raíz cuadrada y elevar el resultado a *N*. Por lo tanto, el precio tras *N* días, siempre en una trayectoria en la que el precio baja en la mitad de los días y sube en la otra mitad, es la potencia *N*-ésima de la *media geométrica* de los dos posibles factores, es decir, $\sqrt{1{,}3 \times 0{,}75} \sim 0{,}987$, que es menor que la unidad. Al elevar a una potencia muy grande un número menor que la unidad, se obtiene una cantidad muy pequeña. Por lo tanto, la fórmula anterior declara que el precio se hace cada vez más pequeño, al menos en una trayectoria "típica", en la que el precio sube el mismo número de días que baja. Pero, ¿importan estas trayectorias "típicas"?

Ya hemos visto en la figura que el precio de la acción de nuestro ejemplo sufre grandes fluctuaciones. Algunas de las posibles trayectorias crecen mucho. Imagínense por ejemplo que, después de 100 días, el precio ha crecido 60 de ellos y ha decrecido 40. El precio final sería:

$$1{,}30^{60} \times 0{,}75^{40} = 664{,}47 \text{ euros}$$

que es muy alto. En la figura vemos también un caso en el que el precio ha subido hasta casi 10.000 euros en unos cien días (*curva roja de la gráfica de la izquierda*). Sin embargo, estas trayectorias son muy poco probables y se hacen aún más raras cuanto más avanza el tiempo, ya que entonces las fracciones de días de subida y bajada se aproximan ambas más a 1/2, haciendo que el precio sea cada vez más cercano a la fórmula que hemos deducido para la trayectoria típica. Por lo tanto, las trayectorias típicas consideradas no sólo son las más probables, sino que, para tiempos muy largos, resulta muy difícil que el precio se aleje mucho de ellas, como vemos en la figura.

Sin embargo, algún lector suspicaz podría argumentar lo siguiente. Si el precio de la acción en un día es *x*, al día siguiente será 1,30*x* con probabilidad 1/2 o 0,75*x* con probabilidad 1/2. Por tanto, el valor medio del capital al día siguiente es 1,025*x*. En otras palabras, el valor medio del capital viene determinado por la media aritmética y no por la geométrica, y además crece día tras día. El precio de la acción es una cantidad aleatoria singular: ¡su valor medio crece, pero es cada vez más improbable que se aleje de cero! ¿Cómo reconciliar dos comportamientos tan opuestos? La respuesta está en esas trayectorias raras, aunque afortunadas. Algunas son extremadamente afortunadas y extremadamente raras, como, por ejemplo, la que resultaría de subir el precio todos los días. A pesar de que ocurren con una probabilidad insignificante, el precio alcanza en ellas un valor tan alto, que influye en la media, haciendo que crezca en lugar de disminuir. Esto ocurre incluso a pesar de que la gran mayoría de las trayectorias da lugar a precios ridículamente pequeños. ¿Con qué deberíamos quedarnos entonces? ¿Con la media o con el comportamiento típico? Creo que ninguno de los lectores lo dudaría, especialmente si se juega su propio dinero: uno espera que la acción siga un comportamiento típico, y la media, que en otros problemas de probabilidad sí desempeña un papel importante, aquí nos da una información distorsionada por las trayectorias afortunadas.

¿Qué ocurre con la inversión diversificada? En este caso, cada día repartimos todo el capital entre las dos acciones. Si tenemos un capital *x*, invertiremos en cada una de las dos acciones una cantidad *x*/2. Pueden ocurrir varias posibilidades. La primera es que las dos acciones suban un 30 %, con lo cual, el capital que tenemos al día siguiente es:

$$\frac{x}{2} \times 1{,}30 + \frac{x}{2} \times 1{,}30 = 1{,}30x.$$

Esta posibilidad ocurre con una probabilidad 1/4. Puede que ambas acciones bajen un 25 %, algo que ocurre también con una probabilidad 1/4 y tras lo cual el capital se reduce a 0,75*x*. Finalmente, puede que una suba y la otra baje. En este caso, el capital pasa a ser

$$\frac{x}{2} \times 1{,}30 + \frac{x}{2} \times 0{,}75 = 1{,}025x.$$

Esta última posibilidad ocurre con una probabilidad 1/2, ya que da igual cuál de las dos suba y cuál baje. Repitiendo el argumento que elaboramos para el caso de una única acción, es fácil deducir que el capital después de *N* días es

$$x_N = 1{,}3^{n_{ss}} \times 0{,}75^{n_{bb}} \times 1{,}025^{n_{sb}}$$

en donde n_{ss} es el número de días en los que las dos acciones han subido, n_{bb} el número de días en los que ambas han bajado y n_{sb} el número de días en los que una ha subido y la otra ha bajado. Teniendo en cuenta las probabilidades respectivas de cada una de estas posibilidades, para un número *N* muy grande de días, una trayectoria "típica" de nuestra inversión da lugar al siguiente capital:

$$x_N = 1{,}3^{N/4} \times 0{,}75^{N/4} \times 1{,}025^{N/2} = \left(\sqrt[4]{1{,}3 \times 0{,}75} \times \sqrt{1{,}025}\right)^N = 1{,}006^N$$

y ahora el número que se eleva a *N* es mayor que uno, con lo cual, el capital crece indefinidamente. Observen que la media aritmética del factor por el que se multiplica el capital es la misma tanto para la inversión en una sola acción como para la inversión diversificada (en este segundo caso es 1,30/4 + 0,75/4 + 1,025/2 = 1,025). Es decir, el valor medio del capital es el mismo tanto si invertimos nuestro dinero en una acción como si diversificamos la inversión. Sin embargo, la media geométrica, que, como hemos visto, es la importante para las trayectorias típicas, ha cambiado a nuestro favor.

¿Cómo es posible que la combinación de dos inversiones perdedoras dé lugar a una ganadora? Cuando repartimos el capital entre las dos acciones estamos promediando las ganancias conseguidas; el capital en una realización típica se acerca más a su valor medio. Si se combinan más de dos inversiones, el resultado es aún mejor. Pero conténganse a la hora de aplicar esta estrategia en bolsa. Aquí hemos partido del supuesto de que las dos acciones que se combinan evolucionan de forma independiente. Y en el mundo real hay fuertes correlaciones entre unos valores y otros. Especialmente cuando bajan todas las cotizaciones al unísono en una crisis financiera.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Hagan sus apuestas

Como se supone que soy un experto en juegos de azar, mucha gente me pregunta si no he aprovechado tanta sabiduría para hacerme rico en un casino. Lo cierto es que la única vez que he jugado a la ruleta fue hace muchos años, fascinado por un método de apuestas supuestamente infalible pero con el que perdí casi todos mis ahorros en menos de un cuarto de hora.

Existe un método clásico, probablemente conocido por muchos lectores, con el que, en teoría, siempre se puede ganar en la ruleta. Casanova lo describe en sus memorias, publicadas en 1754. Se apuesta siempre a rojo, a negro o a cualquier jugada que dé como premio la misma cantidad que se apuesta. Se comienza con una apuesta baja, un euro, por ejemplo. Cada vez que se pierde, se dobla la cantidad apostada, mientras que si se gana se vuelve a empezar el proceso con un euro. Si, por ejemplo, perdemos cinco veces seguidas, las apuestas que debemos realizar son 1, 2, 4, 8 y 16 euros en cada turno, lo que hace un total de 31 euros perdidos. En el sexto turno apostaremos 32 euros y por tanto, si ganamos, recuperamos todo lo perdido más un euro. Ocurre lo mismo con cualquier número de turnos, debido a una conocida fórmula matemática:

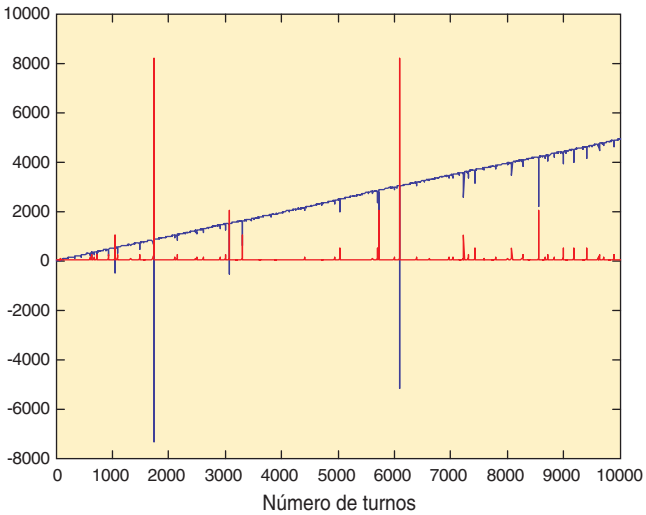
$$1 + 2 + 2^2 + 2^3 + \ldots 2^n = 2^{n+1} - 1$$

Con este sistema basta ganar una vez para ganar un euro, sin importar cuántas veces hayamos perdido antes de ese turno ganador. Como es imposible perder un número infinito de veces, con el sistema de Casanova, que se conoce como *martingala*, se ganaría siempre. ¿Por qué entonces no se ha arruinado todavía ningún casino? La respuesta es muy simple: porque es bastante probable perder un gran número de veces seguidas y en ese caso la apuesta puede superar la apuesta máxima permitida o, lo que es peor, nuestros propios fondos.

En un conocido casino de Madrid, la apuesta máxima está entre 20 y 30 veces la apuesta mínima, con lo que la martingala alcanzaría la apuesta máxima con sólo perder cinco veces seguidas y esto ocurre con una probabilidad $(19/37)^5 = 0{,}03507$ (hay 37 números en la ruleta, de los que, si apostamos a rojo o a negro, 19 son perdedores dado que el cero es siempre perdedor).

Incluso con apuestas máximas más altas, la martingala entraña un considerable riesgo. En la figura 1 pueden ver la evolución del capital (*en azul*) y de la cantidad apostada (*en rojo*) de un jugador que aplica el método de martingala durante diez mil turnos, comenzando con un euro. Como vemos, el jugador gana de forma sistemática. Como gana un euro por cada turno ganador, la ganancia total es, aproximadamente, igual a la mitad del número de turnos. En la gráfica se aprecia que el capital se encuentra cerca de una recta de pendiente 1/2. Sin embargo, hay ciertos periodos de mala suerte en donde un número continuado de pérdidas hace que el capital baje considerablemente. Vemos una de estas malas rachas entre el turno 1000 y 2000, en donde la apuesta sube hasta 8192 (*curva roja*), después de 12 turnos sin ganar, y el capital desciende hasta casi –8000 euros. Es decir, nuestro jugador habría necesitado un "colchón" de unos 8000 euros para superar esta mala racha. Lo malo del método de martingala es que estas malas rachas pueden ser de cualquier intensidad. En la gráfica vemos que ha habido dos apuestas de 8192 euros, que es la cantidad a apostar después de 12 turnos sin ganar. Pero bien podría haber sido también perdedor el turno decimotercero, con lo que la apuesta habría subido a más de 16.000 euros. En definitiva, la martingala es un método muy arriesgado. De hecho, algunos casinos han aumentado considerablemente las apuestas máximas ante la evidencia de que nadie se atreve a poner en práctica la martingala, al menos en su formulación original. En algunos casinos de Las Vegas, la apuesta máxima puede llegar a ser 4.000 veces la apuesta mínima. En ellos la martingala alcanzaría la apuesta máxima tras perder 11 veces seguidas. Podría parecer que el casino está arriesgándose excesivamente. Sin embargo, para que un jugador gane, por ejemplo, 1000 euros, necesitaría jugar aproximadamente 2000 veces y en esas dos mil veces la probabilidad de perder 11 veces seguidas es de más de un 50 %.

En algunos libros y sitios de Internet se aconseja esperar a que, por ejemplo, salga rojo 5 veces seguidas para comenzar a utilizar la martingala apostando a negro, con la supersticiosa creencia de que es prácticamente imposible que en una tarde el rojo salga 15



1. Capital (*en azul*) y apuesta realizada (*en rojo*) en 10.000 turnos utilizando el método de la martingala.

veces seguidas. Cualquier experto en probabilidad sabe que una ruleta no tiene (o no debe tener) memoria alguna. Aunque haya salido rojo 100 veces seguidas, la probabilidad de que salga rojo o negro en la siguiente tirada es exactamente la misma.

Hay otros métodos inspirados en la martingala pero, en apariencia, más razonables. Uno de ellos es el que consiguió engatusarme hace años, hasta que la caprichosa bolita de la ruleta me devolvió a la dura realidad. Se conoce como método de Labouchere, en honor de Henry du Pre Labouchere, un periodista y parlamentario inglés que vivió entre 1831 y 1912. Sin embargo, el propio Labouchere atribuía el método al matemático francés Condorcet (1743-1794), de quien ya hemos hablado en esta sección (*Paradojas democráticas*, marzo de 2002). El método consiste en escribir en un papel una lista con unos pocos números, por ejemplo, 1 1 3. En cada turno apostamos a rojo o negro la suma del primero y el último número de la lista. En nuestro caso, la primera apuesta sería de 4 euros. Cada vez que ganamos, tachamos el primero y el último número, mientras que, si perdemos, añadimos a la lista la cantidad perdida. Cuando se han tachado todos los números de la lista se vuelve a empezar. La siguiente tabla muestra cómo evoluciona el capital (empezando con cero euros), la apuesta y la lista en unos cuantos turnos:

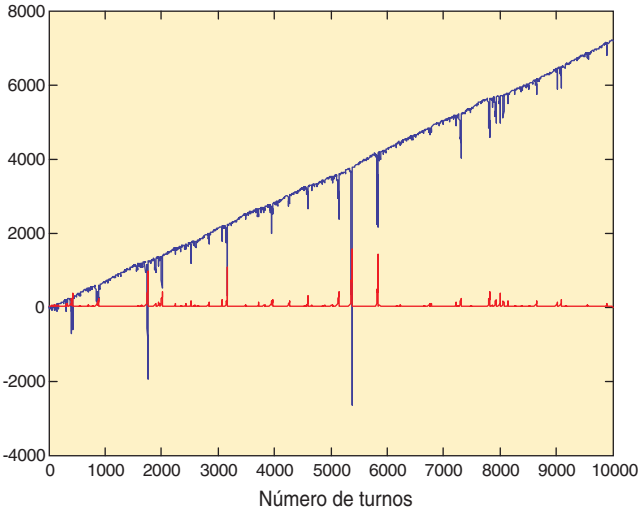
LISTA	APUESTA	RESULTADO	CAPITAL
1 1 3	4	P	-4
1 1 3 4	5	P	-9
1 1 3 4 5	6	G	-3
1 3 4	5	P	-8
1 3 4 5	6	G	-2
3 4	7	P	-9
3 4 7	10	P	-19
3 4 7 10	13	G	-6
4 7	11	G	+5

Después de estos 9 turnos, hemos tachado todos los números de la lista. Puesto que todo lo que se gana se tacha de la lista y todo lo que se pierde se apunta, la ganancia neta al tachar todos los números es igual a la suma de los números con los que comenzamos la lista, es decir, 1 + 1 + 3 = 5 euros. Pero cada vez que ganamos tachamos dos números (salvo si en la lista sólo queda un número) mientras que apuntamos sólo uno cada vez que perdemos. Por lo tanto, si perdemos n veces, bastará ganar $(n + 3)/2$ veces (o el entero inmediatamente superior si n es par) para tachar todos los números y ganar 5 euros. En el ejemplo de la tabla hemos perdido 5 veces y hemos ganado 4. Esta diferencia se hace mayor cuantos más turnos juguemos antes de tachar todos los números. Si, por ejemplo, perdemos 20 veces, basta ganar 12 para conseguir los 5 euros. El método de Labouchere se basa en la misma idea que la martingala: aumentar la apuesta cuando se pierde para, de ese modo, tener que ganar menos veces para recuperar lo perdido. La ventaja frente a la martingala es que la apuesta crece mucho más lentamente, con lo cual parece difícil alcanzar la apuesta máxima del casino. Sin embargo, esta apreciación no se corresponde con la verdad. En la figura 2 se muestra una simulación del método y vemos que en ocasiones la apuesta (*curva roja*) casi alcanza los 2000 euros y que el capital desciende hasta casi --3000. El método de Labouchere es especialmente peligroso ante rachas en las que se pierden cuatro o cinco turnos seguidos y a continuación se gana sólo una vez. En ese caso, la lista se va poblando de números cada vez más altos.

En la figura 2 vemos que la ganancia media en 10000 turnos es del orden de 7000 euros. Calcular la ganancia media de forma exacta es un problema matemático bastante difícil. Sin embargo, se puede estimar de una forma muy sencilla. Como hemos visto antes, con el método de Labouchere, si se pierde durante n turnos, basta ganar durante los $(n + 3)/2$ siguientes (o el entero inmediatamente superior) para ganar 5 euros. Podemos escribir estos números en una tabla:

TURNOS PERDIDOS	1	2	3	4	5	6	7	8	9
TURNOS GANADOS	2	3	3	4	4	5	5	6	6
GANANCIA	+5	+5	+5	+5	+5	+5	+5	+5	+5

Si jugamos un gran número de turnos, ganaremos en la mitad de ellos y perderemos en la otra mitad. Las combinaciones más comunes de turnos ganados y perdidos son entonces las resaltadas en azul en la tabla. Por tanto, lo más probable es que ganemos 5 euros cada 7 turnos. En el caso de 10000 turnos, la ganancia tiene que ascender a $5 \times 10.000/7 \approx 7143$ euros, de acuerdo con la simulación de la figura 2.



2. Capital (*en azul*) y apuesta realizada (*en rojo*) en 10.000 turnos utilizando el método de Labouchere.

Espero que se tomen estos métodos como meras curiosidades matemáticas y no los apliquen en el casino. Para olvidar semejante idea, basta echar un vistazo a las dos figuras e imaginarse a uno mismo aumentando más y más la apuesta en una de esas malas rachas. Eso es precisamente lo que me ocurrió hace años, utilizando el método de Labouchere. La apuesta subió tanto, que estuve a punto de perderlo todo al poco tiempo de empezar a jugar.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Quien ríe el último...

En el último libro del matemático John Haigh, *Matemáticas y Juegos de Azar*, hay una buena colección de paradojas y curiosidades sobre probabilidad y de análisis interesantes de conocidos concursos de televisión, loterías y quinielas.

Uno de los ejemplos más sorprendentes es el llamado *Penney-ante*, un juego propuesto por Walter Penney en 1974. Se juega con una moneda completamente equitativa, es decir, una moneda en la que los dos resultados al lanzarla, cara o cruz, tienen la misma probabilidad, 1/2. El juego es muy simple: cada jugador escoge una secuencia de tres tiradas, por ejemplo CARA-CRUZ-CARA; se lanza la moneda repetidas veces y gana el jugador cuya secuencia aparece en primer lugar.

Supongamos que jugamos usted y yo al Penney-ante. En un alarde de generosidad, yo le ofrezco elegir en primer lugar su secuencia. Puede escoger cualquiera de las ocho secuencias:

CCC CCE CEC CEE
ECC ECE EEC EEE

Utilizo C para Cara y E para Cruz o Escudo. Una vez que usted haya hecho su elección, yo elijo mi secuencia y comenzamos a lanzar la moneda. A primera vista, elegir en primer lugar sólo puede reportar ventajas: si hubiera alguna secuencia mejor que el resto, puede elegirla y dejarme a mí una menos favorable; si todas las secuencias son igualmente probables, da igual quién elige primero. Sin embargo, en este juego ocurre algo bastante curioso: para cualquier elección suya, yo puedo elegir una secuencia mejor, es decir, una secuencia que aparezca antes que la suya con una probabilidad mayor que 1/2. ¡Elegir primero resulta ser una desventaja! El Penney-ante es parecido al conocido juego de piedra, papel y tijeras, en el que, si un jugador eligiera primero y mostrara la elección al contrincante, perdería con toda seguridad.

En algunos casos la segunda elección es bastante sencilla. Si, por ejemplo, usted elige la secuencia CCC, entonces mi elección es bastante fácil: ECC. La única forma de que salga CCC antes que CCE es que el resultado de los tres primeros lanzamientos sea cara, lo que ocurre sólo con una probabilidad 1/8. Si la secuencia CCC aparece *por primera vez* en algún momento que no sea en las tres primeras tiradas, está claro que la tirada anterior a la primera de las tres caras será necesariamente cruz y, por tanto, ECC habrá aparecido antes.

Supongamos que usted elige ECC. En este caso, puede demostrarse que, si yo elijo EEC, tendré una probabilidad de ganar mayor que la suya. El diagrama de la figura explica el origen de esta ventaja. En cada una de las casillas del diagrama se muestran los resultados de las últimas tiradas y las flechas corresponden a los posibles resultados de la tirada siguiente (cara, flechas azules; cruz, flechas amarillas). Si no ha aparecido ninguna de las secuencias en juego y el último resultado es cara, es como si empezáramos el juego, puesto que ninguna de las secuencias comienza por cara. El resto del diagrama es evidente. Por ejemplo, si estamos en la casilla... E y el resultado de la tirada es E, saltamos a la casilla... EE. Si sale de nuevo una E, volvemos a la casilla... EE. Se han especificado sólo las secuencias relevantes para el juego. Podríamos haber añadido una casilla... EEE; mas, para los efectos del juego con las secuencias elegidas, si las dos últimas tiradas son cruz, la antepenúltima es irrelevante.

Según los resultados de las tiradas, nos moveremos por las casillas del diagrama hasta alcanzar la verde, en cuyo caso gano yo, o la roja, en cuyo caso gana usted. Pero el diagrama no es simétrico. Una vez que han aparecido dos cruces seguidas (... EE), no hay forma de volver hacia atrás. En otras, palabras, ganaré con toda seguridad la apuesta. Sin embargo, la secuencia... EC no le asegura a usted la victoria, puesto que una cruz hace que el juego vuelva al estado... E. Cada vez que nos encontramos en el estado... E, la secuencia EEC tiene una probabilidad 1/2 de ganar, mientras que la ECC sólo tiene una probabilidad 1/4, puesto que necesita dos caras seguidas. Por lo tanto, la probabilidad de que yo gane es el doble de la suya, es decir, yo gano con una probabilidad 2/3 y usted con una probabilidad 1/3.

Acabamos de ver que EEC es claramente mejor que ECC. Por lo tanto, en el siguiente turno usted puede elegir EEC. Sin embargo, en este caso yo tengo de nuevo una elección ganadora. Si elijo CEE, también

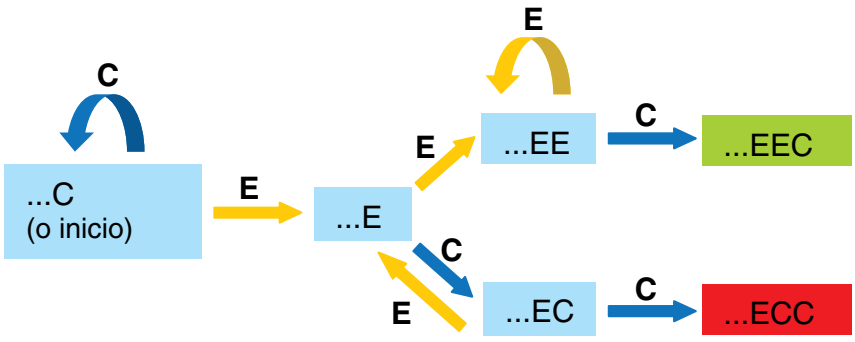


Diagrama del Penney-ante. El juego se mueve por las casillas hasta caer en la roja (usted gana) o en la verde (yo gano).

tendré una probabilidad doble que la suya de ganar. Lo sorprendente es que, ante cualquier elección suya, yo puedo elegir una secuencia que me dé una ventaja de al menos 2/3. En la siguiente tabla se muestran todas las posibilidades:

Para calcular la probabilidad de ganar cada uno de los casos, tenemos que dibujar un diagrama ligeramente diferente del de la figura. Habrá que escoger las secuencias relevantes para cada fila y dibujar las flechas correspondientes, algo no excesivamente difícil a la luz del ejemplo ECC-EEC de la figura.

Su elección	Mi elección	Probabilidad de que yo gane
CCC	ECC	7/8
CCE	ECC	3/4
CEC	CCE	2/3
CEE	CCE	2/3
ECC	EEC	2/3
ECE	EEC	2/3
EEC	CEE	3/4
EEE	CEE	7/8

Por otro lado, la tabla obedece a una regla bastante simple que permite obtener una secuencia que venza a cualquier otra: tome los dos primeros elementos de la secuencia elegida en primer lugar, colóquelos al final y elija el primer elemento de la nueva secuencia de modo que el resultado no sea capicúa. Por ejemplo, para averiguar la secuencia que vence a CCE, tomamos CC y elegimos como primer elemento una E para formar ECC. La regla pone en evidencia que la secuencia ganadora se "aprovecha" de parte de la secuencia perdedora.

Es evidente que no hay ninguna secuencia óptima. De hecho, observamos una circularidad similar a la de piedra, papel y tijeras: CCE vence a CEE, que a su vez vence a EEC, que a su vez vence a ECC, que vence a CCE. En esta serie de secuencias, cada una comienza por donde termina la anterior, como impone la regla que hemos discutido antes:

C C E
C E E
E E C
E C C
C C E

El juego y la regla para obtener secuencias ganadoras se pueden generalizar a secuencias de más de tres tiradas. Cuando una relación como "vencer a" tiene este tipo de circularidad, se dice que es "no transitiva". Por el contrario, una relación *r* es transitiva cuando, si A*r*B y B*r*C, entonces necesariamente A*r*C. La falta de transitividad en ciertas relaciones se opone muchas veces a la intuición. El propio juego de piedra, papel y tijeras fascina a los niños por su no transitividad. En competiciones deportivas, nos cuesta aceptar que A gane a B, B a C, pero que sea C quien gane a A. Por mucho que veamos este tipo de situaciones, siempre tendemos a pensar que "algo le ha pasado" al equipo que supuestamente debería ganar. En esta sección apareció un ejemplo también sorprendente de no transitividad: la paradoja de Condorcet, en la que las preferencias de un grupo de votantes ante tres candidatos no son transitivas (*Paradojas democráticas*, marzo 2002). Es decir, puede darse el caso de una elección en donde el candidato A gane a B en una votación donde sólo se vote por uno de los dos, B gane a C y C gane a A.

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Finalmente... sudoku

Era inevitable. No podía pasar un mes más sin que el famoso sudoku apareciera en estas páginas. El "pasatiempo del verano" ha sido sin duda este sencillo rompecabezas de origen japonés que lleva ya varios años causando furor en el Reino Unido y que en pocos meses se ha hecho un hueco en las páginas de casi todos los periódicos de nuestro país. Podemos recibir sudokus incluso en el teléfono móvil.

El sudoku, por si alguien aún no lo sabe, consiste en rellenar un tablero cuadrado de 9×9 casillas, como el de la figura, con números del 1 al 9, de modo que no se repita ningún número en cada fila, en cada columna y en cada uno de los 9 bloques de 3×3 casillas delimitados por líneas gruesas. Cada sudoku se presenta con algunas cifras ya colocadas que determinan de forma unívoca el resto del tablero. Aunque se utilicen números para rellenar las casillas, el sudoku no tiene nada que ver con la aritmética; se trata de un rompecabezas de pura lógica: en lugar de números podrían utilizarse letras o colores. La clave para resolverlo es encontrar casillas en donde necesariamente debe escribirse un número determinado. En el sudoku de la figura 1, por ejemplo, el 8 del bloque inferior-izquierda debe estar en la casilla central, ya que el resto de filas y columnas tienen ya un 8. Esta es una deducción no muy complicada. En ocasiones hay que tener en cuenta más posibilidades e incompatibilidades hasta dar con el número que necesariamente se debe colocar en una cierta casilla. Cada sudoku tiene su nivel de dificultad —desde "muy fácil" hasta "diabólico"—; suele ser más alto cuantos menos números se presentan inicialmente (aunque no siempre es así).

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

A pesar de su sencillez, el sudoku plantea algunos problemas matemáticos y lógicos bastante complicados: ¿cuántos sudokus diferentes se pueden construir? ¿Existen algoritmos para resolver sudokus de forma automática? ¿De qué depende el nivel de dificultad de un sudoku? ¿Cuál es el mínimo número de cifras iniciales que determinan unívocamente el sudoku?

La primera de estas preguntas ha sido resuelta esta pasada primavera por Bertram Felgenhauer, de la Universidad Técnica de Dresde, y Frazer Jarvis, de la Universidad de Sheffield, con ordenadores personales y una buena dosis de lógica y combinatoria. El resultado es bastante sorprendente. El número de sudokus posibles es $6.670.903.752.021.072.936.960 \approx 6{,}671 \times 10^{21}$. Es sorprendente porque puede escribirse como $(9!) \times \times 72^2 \times 2^7 \times 27.704.267.971$. Cada uno de estos factores puede explicarse razonablemente salvo el último de ellos, un número primo de tamaño considerable obtenido por Felgenhauer el 23 de mayo de 2005, tras seis horas de computación en dos ordenadores personales. El programa informático de Felgenhauer fue el fruto de una colaboración entre varias personas a través de un foro en Internet (www.sudoku.com/forums/viewtopic.php?t=44). Es fascinante seguir paso a paso las discusiones que mantuvieron durante días, en las que se conocieron Frazer y Bertram, hasta dar con la solución final. Aunque encontrar el número de sudokus no sea una gran proeza matemática, sí creo que es un caso único de que hayan quedado registrados de manera exhaustiva los avatares de una colaboración científica, desde el planteamiento hasta la solución final del problema. Un registro además muy humano, porque los mensajes están repletos de intentos fallidos, dudas y entusiasmo.

Encontrar el número de sudokus por el método de la fuerza bruta, es decir, pidiendo a un ordenador que rellene de todas las formas posibles las 81 casillas, es una estrategia inviable. El ordenador más potente tardaría años en terminar esta tarea. Felgenhauer y Jarvis tuvieron que simplificar el cálculo buscando simetrías en el problema hasta lograr que la búsqueda se redujera a unas cuantas horas. Veamos alguno de estos trucos. En primer lugar, no es difícil deducir el número de combinaciones posibles para los tres bloques superiores del tablero. Conviene denotar cada uno de los nueve bloques del sudoku de la siguiente forma:

B1	B2	B3
B4	B5	B6
B7	B8	B9

en donde cada bloque es un cuadrado de 3×3 casillas. El primer bloque, B1, se puede llenar de $9! = 362.880$ maneras diferentes, porque $9!$ es el número de posibles reordenaciones de 9 elementos (en este caso, las nueve cifras). Supongamos la siguiente configuración para B1:

1	2	3
4	5	6
7	8	9

y veamos las posibles formas de rellenar B2. Los números 4, 5 y 6 tienen que estar en la fila superior o inferior de B2. Supongamos que llevamos los tres números a la fila superior. En este caso, 7, 8 y 9 deberán hallarse necesariamente en la fila central y 1, 2 y 3 en la fila inferior.

Si, por el momento, no nos ocupamos de la columna en la que colocamos cada número, concluimos: si 4, 5 y 6 están en la fila superior o inferior de B2, el resto de filas quedará completamente determinado. Por lo tanto, estas dos opciones dan lugar a 2 posibles configuraciones en B2 (siempre sin tener en cuenta las ordenaciones de los números en cada fila). Si en B2 colocamos el 4 en la fila inferior y el 5 y el 6 en la superior, obtendremos más configuraciones posibles:

B1			B2		
1	2	3		5	6
4	5	6			
7	8	9	4		

En la fila central superior debemos colocar ahora una cifra del conjunto {7, 8, 9} y en la inferior dos cifras del conjunto {1, 2, 3}. Esto lo podemos hacer de $3 \times 3 = 9$ formas distintas. Lo mismo ocurre si colocamos el 5 en la fila superior y el 4 y el 6 en la inferior. Hay $3 \times 2 = 6$ configuraciones de este tipo: dos elementos de {4, 5, 6} arriba y uno abajo, o viceversa. Por lo tanto, y siempre sin fijarnos en el orden horizontal de los números, tenemos $2 + 9 \times 6 = 56$ formas de colocar los números en B2. Pero si consideramos que los tres números de cada fila pueden ordenarse de $3 \times 2 = 6$ maneras distintas, llegamos a que el número de combinaciones posibles para B2 es $56 \times 6^3 = 12.096$.

Finalmente, si B1 y B2 están llenos, los números de cada fila en B3 quedarán completamente determinados. Tenemos de nuevo 6^3 ordenaciones de las tres filas. Por lo tanto, el número total de disposiciones para B1, B2 y B3 es $(9!) \times 56 \times 6^3 \times 6^3 = 948.109.639.680$.

Un cálculo de fuerza bruta supondría tomar cada una de estas disposiciones y probar con ella todas las combinaciones posibles para el resto de los bloques B4-B9. Sin embargo, este cálculo sigue siendo inviable. La idea de Jarvis y Felgenhauer es reducir las combinaciones B1-B3 a ciertas clases. Dos combinaciones B1-B3 están en la misma clase si tienen el mismo número de combinaciones B4-B9 compatibles. Apoyados en argumentos de simetría, redujeron el número de clases a 71. Detallar todos estos argumentos de simetría excede el espacio de esta sección. Algunos son muy simples. Por ejemplo, cualquier reetiquetado de los números en B1-B3 da lugar a una combinación de la misma clase. Como hay 9! reetiquetados, el número de combinaciones a estudiar se reduce a $56 \times 6^3 \times 6^3 = 2.612.736$. Otras permutaciones de los números, casillas o ambos en B1-B3 reducen el número de clases a 71. De hecho, después de ejecutar el programa, Jarvis y Felgenhauer comprobaron que en realidad hay sólo 44 clases, una reducción que probablemente se deba a simetrías en las que ellos mismos no habían reparado. El lector interesado puede encontrar los detalles buscando "Jarvis" y "sudoku" en Google, para acceder a la página web en donde los autores han colocado una descripción exhaustiva de sus resultados y del método utilizado.

Otro participante en la discusión de la que hablábamos, Kevin Kilfoil, había llegado a un resultado bastante aproximado con un argumento sencillo y muy elegante. El argumento es el siguiente. Supongamos que sólo consideramos la regla "no pueden repetirse números en cada bloque". En ese caso, el número de posibles configuraciones es $(9!)^9 \approx 1{,}0911 \times 10^{50}$ ya que cada uno de los bloques admite 9! configuraciones. Si ahora consideramos la regla "no pueden repetirse números en cada fila", el número de configuraciones posibles se reduce. Hemos visto que con esta regla, los bloques B1-B3 se pueden rellenar de $(9!) \times 56 \times 6^3 \times 6^3$ maneras distintas. Lo mismo ocurre para los bloques B4-B6 y B7-B9. Por tanto, con las dos reglas, la de los bloques y la de las filas, el número de configuraciones posibles es $[(9!) \times \times 56 \times 6^3 \times 6^3]^3$. La regla de las filas reduce entonces el número de configuraciones en un factor:

$$R = \frac{(9!)^9}{(9! \times 56 \times 6^6)^9} = \frac{(9!)^9}{56^3 \times 6^{18}} \approx 1{,}28 \times 10^{14}$$

La idea ingeniosa de Kilfoil es suponer que la regla "no pueden repetirse números en cada columna" actúa de forma similar, reduciendo el número de combinaciones por el mismo factor R. El número total de configuraciones de sudoku posibles sería entonces:

$$\frac{(9!)^9}{R^2} \approx 6657 \times 10^{21}$$

El resultado final no es exacto. De hecho, no es ni siquiera un número entero, pero difiere de la solución exacta en sólo un 0,2 %. Es curioso que la hipótesis de que la regla de las filas y la regla de las columnas actúan de forma independiente dé lugar a una estimación tan correcta. Estamos ante un ejemplo de cómo atacar un problema combinatorio con argumentos aproximados y en cierto modo probabilísticos (podríamos reformular el argumento de Kilfoil suponiendo que la "probabilidad" de que una configuración que respeta la regla de los bloques respete también la de las filas o la de las columnas es 1/R y que estas probabilidades son independientes). Algo parecido a lo que vimos hace unos meses con los primos aleatorios de Hawkins [*véase* "Cribas y números primos", agosto 2005].

Quedan otras muchas cuestiones acerca de la combinatoria y matemática del sudoku. Nos preguntábamos al comienzo de este artículo si existe un algoritmo para solucionar cualquier sudoku. La respuesta es que existen varios: desde el cálculo por la "fuerza bruta", hasta la implementación en un programa informático de las mismas estrategias que utilizan los adictos al sudoku. En http://sudoku.sourceforge.net/ pueden encontrar un algoritmo de este tipo. Se trata de un programa Java de muy fácil uso que resuelve cualquier sudoku y explica las reglas que ha aplicado en la solución. Este programa permite saber además si un sudoku con algunos números ya colocados en ciertas casillas admite solución y si ésta es única. Es interesante jugar con el programa introduciendo números hasta llegar a sudokus con solución única, algo que no es fácil de conseguir.

Otra de nuestras preguntas iniciales es el mínimo número de casillas inicialmente llenas para un sudoku con solución única. Parece que es 17, pero nadie ha podido probarlo todavía. Una discusión muy amplia sobre este tema se puede encontrar también en el foro www.sudoku.com/forums/viewtopic.php?t=605.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

¿Hay quien dé más?

Imagine la siguiente lotería. Usted elige un número del 1 al 100 y lo anota en un papel. A continuación se extrae un número de un bombo con 100 bolas, numeradas de 1 a 100. Usted gana si el número que anotó es *menor o igual* que el número que sale del bombo, pero gana la cantidad anotada en euros. Si el número en el papel es mayor que el que ha salido en el sorteo, entonces usted no gana nada.

¿Cuál es la mejor elección inicial? Si elegimos un número bajo, la probabilidad de ganar es mayor pero la ganancia será pequeña. Por el contrario, al elegir un número grande aumentamos la ganancia, pero reducimos las probabilidades de ganar. Llamemos M al número elegido inicialmente. La probabilidad de ganar es $(101-M)/100$, puesto que, de los 100 números que hay en el bombo, $101-M$ son mayores o iguales que M. La ganancia media al elegir M es por tanto:

$$\frac{101-M}{100}\times M$$

Como cabría esperar, si elegimos $M = 1$, la ganancia media es de un euro. Si elegimos $M = 100$, también la ganancia media es de un euro, porque ganamos 100 euros pero sólo un 1 % de las veces. El número M que hace mayor la ganancia media es $M = 50$ o 51. En ambos casos, la ganancia media es $50\times 51/100 = 25{,}5$.

Si consideramos una variante de la lotería anterior en la que lo que usted gana es la cantidad que ha aparecido en el bombo, en lugar de la que anotó en el papel, entonces la estrategia sería bastante más sencilla. Puesto que en este caso la ganancia no depende de lo que ponga usted en el papel, la mejor elección es $M = 1$. Con ello siempre ganamos y la ganancia media será de 50,5, que es el valor medio de todos los números del bombo.

Pero el título de la columna de hoy hace referencia a una subasta. ¿Qué tienen que ver estas extrañas loterías con una subasta? Cuando participamos en ciertos tipos de subasta, nos encontramos en una situación similar, aunque inversa, al primero de los sorteos que acabamos de discutir. Si pujamos por una cantidad demasiado alta, probablemente ganemos la subasta pero pagaremos un precio excesivo por el artículo subastado. Si queremos pagar menos, tendremos que pujar con cantidades menores disminuyendo así la probabilidad de adquirir el artículo. Hay un compromiso obvio entre el deseo de ganar la subasta y el precio a pagar. Esta es la principal característica de una subasta y, en cierto modo, su razón de ser. Sin embargo, a pesar de la aparente sencillez de estos mecanismos, su análisis matemático puede depararnos alguna que otra sorpresa. El estudio de las subastas que el norteamericano William Vickrey realizó en los años sesenta, y su posterior extensión a otras situaciones, le hizo merecedor, en 1996 y junto con James A. Mirrlees, del Premio del Banco de Suecia, el mal llamado Nobel de economía.

En primer lugar, existen varios mecanismos de subasta. En la subasta inglesa o de oferta creciente, los participantes, que están presentes en el proceso de subasta, anuncian públicamente sus pujas en orden ascendente y gana la mayor. En las lonjas de nuestros puertos se lleva a cabo un sistema diferente y bastante espectacular, la subasta de precio decreciente. En ella el director va cantando precios en orden descendente hasta que alguno de los participantes le interrumpe gritando "¡mío!", con lo que consigue el artículo subastado por el último precio cantado. Se llama también holandesa porque era éste el método que utilizaban desde el siglo XVII los mayoristas de flores en Holanda para distribuir sus existencias. ¿Cuál de los dos métodos es mejor para el subastador y para el comprador? La subasta de precio descendente es bastante más estresante para el comprador. Al fin y al cabo, en la subasta de oferta ascendente uno sabe siempre que puede quedarse con el artículo subastado si supera la puja de los demás. Sin embargo, en la de precio descendente, hay una incertidumbre considerable: si espero para conseguir un mejor precio me arriesgo a que otro gane la subasta; si me planto en un precio alto quizá haya perdido la oportunidad de ahorrarme una buena cantidad.

En muchas ocasiones se necesita realizar una subasta sin la presencia física de los participantes. Si, por ejemplo, queremos que el número de participantes sea el máximo posible, es mejor utilizar algún método en el que se pueda pujar "a distancia". Ocurre así en las subastas de Internet, cada vez más populares. Pero ya en el siglo XIX se realizaban subastas de sellos de correo en donde se pujaba por carta. Es también el caso de algunas grandes operaciones financieras o de las adjudicaciones de licencias de telefonía móvil, que en muchos países (pero no en España) se realizaron mediante subastas de este tipo. No sé cuál es la razón de no usar en estos casos subastas con presencia de los licitadores. Quizá sea porque da mala impresión que una gran empresa decida gran parte de su futuro en el calor de una puja pública.

En estas subastas, llamadas de oferta sellada, las pujas se envían en un sobre cerrado al director de la subasta quien, una vez recibidas todas las ofertas o pujas, decide el resultado. Lo más común es que gane la subasta quien ha realizado la oferta más alta y que pague por el artículo subastado la cantidad que ofertó. Este sistema se llama "de primer precio".

Vickrey mostró que la subasta de oferta sellada de primer precio es equivalente a la de precio descendente. No es difícil entender por qué. Supongamos que hacemos un simulacro de subasta de precio descendente con las ofertas recibidas. Cada oferta es el precio que está dispuesto a pagar quien la remitió. Por lo tanto, si comenzamos a cantar precios en orden descendente, el primero en gritar "¡mío!" será quien envió la oferta más

alta. Ganará la subasta y adquirirá el artículo subastado por el precio ofertado, es decir, el precio final será el de la oferta más alta.

¿Tiene la subasta más habitual, la inglesa o de oferta ascendente, su equivalente de oferta sellada? Volvamos a imaginar un simulacro, esta vez de subasta inglesa, realizado con las ofertas recibidas. Imaginemos que cada oferta tiene un "representante" dispuesto a pujar tan alto como indique su oferta correspondiente. Comienza la subasta y las pujas van subiendo, digamos de euro en euro,... ¿hasta la oferta más alta? No. Supongamos que la oferta más alta, llamémosla A, es de 1000 euros y que la *segunda* oferta más alta, B, es de 900 euros. Cuando la puja llega a 900, el representante de A será el único capaz de superarla pujando 901 euros. Pero ahora nadie subirá a 902. Por lo tanto, A gana la subasta y sólo tiene que pagar por el artículo el precio ofertado por B (más la pequeña cantidad de un euro o mínimo incremento de puja), es decir, la *segunda* oferta más alta. Esta subasta de oferta sellada, en la que gana el que ofrece más pero paga no lo que ofertó, sino la segunda oferta más alta, se llama "subasta de Vickrey" o "de segundo precio". A primera vista, es una variante un tanto extraña porque parece que el subastador está tirando piedras contra su propio tejado: si A estaba dispuesto a pagar 1000 euros, ¿por qué venderle el artículo por 900? Vickrey, sin embargo, demostró que tiene virtudes considerables, utilizando un argumento estrechamente relacionado con nuestros sorteos iniciales.

Supongamos que se subasta un objeto por el que está usted dispuesto a pagar como máximo 1000 euros. Tenemos primero que precisar qué entendemos por este precio máximo. No es un precio determinado por sus vicisitudes económicas, su presupuesto mensual o su patrimonio. Es decir, no es 1000 euros la cantidad máxima a pagar porque no se pueda permitir un desembolso de 1001 euros, sino, simplemente, porque para usted el objeto subastado no vale 1001 euros. No tiene un deseo especial de poseerlo ni emociones de ningún tipo. Es sólo pura cuantificación del valor del objeto. Puede considerar la subasta como un juego o lotería en donde, si consigue el objeto por un precio p, gana $1000 - p$ euros, y no gana nada si pierde la subasta. Por lo tanto, le es absolutamente indiferente ganar la subasta por un precio $p = 1000$ euros o irse a casa con las manos vacías.

Con esta premisa, que es consecuencia de la llamada "teoría de la utilidad", una de las bases de la economía clásica, podemos analizar los distintos tipos de subasta de oferta sellada. ¿Cuál es su mejor estrategia cuando participa en una subasta de primer precio? Para simplificar, supongamos que la oferta máxima *del resto* de participantes es un número aleatorio entre 900 y 1000 euros. En este caso, la subasta es similar a nuestro primer sorteo. Si su oferta es p^*, la probabilidad de ganar la subasta es $(p^* - 900)/100$. La ganancia media será entonces:

$$p^* = \frac{p + 900}{2}$$

y resulta ser máxima si $p^* = 950$ euros. Esta es por tanto la oferta que debe realizar. Por otro lado, si participara en una subasta de Vickrey, el precio no dependería de su oferta, como en nuestra segunda lotería. Le interesa entonces maximizar la probabilidad de ganar, lo que se consigue ofertando exactamente 1000 euros. La subasta de Vickrey tiene una virtud importante: la estrategia óptima de todos los participantes es "decir la verdad", ofertar lo que para ellos vale el objeto subastado. Así ocurre también en la subasta inglesa. Observen la diferencia con la subasta de "primer precio", en la que decir la verdad no produce ningún beneficio, porque, como hemos dicho, ganar la subasta por 1000 euros es igual que perderla. Hay que pujar un poco por debajo, al igual que los participantes en la subasta holandesa esperan, con nervios de acero, a que el precio esté por debajo de lo que estarían dispuestos a pagar.

¿Qué ocurre con el subastador? Con la subasta de primer precio obtiene ofertas más bajas pero vende el artículo por el precio más alto. Con la de Vickrey, consigue las ofertas máximas, pero vende por la segunda más alta. ¿Cuál de los dos sistemas le reporta un mayor beneficio? Esta pregunta es más difícil de contestar y depende de cuántas personas participen en la subasta. Podemos resolver parcialmente el problema en un caso muy simplificado en el que los precios p de N participantes se encuentran distribuidos al azar entre, por ejemplo, 900 y 1000 euros. Supondremos también que esta información es conocida por todos los participantes. En este caso, la oferta óptima en la subasta de primer precio para un participante que valore el artículo subastado en un precio p se puede encontrar por un argumento similar al anterior y resulta

$$G = (1000 - p^*)\,\frac{p^* - 900}{100}$$

Luego las ofertas en la subasta de primer precio serán N números distribuidos al azar entre 900 y 950. En el caso de la subasta de Vickrey, las ofertas coinciden con los precios: serán por tanto, N números distribuidos al azar entre 900 y 1000. Para comparar los precios finales en las dos subastas necesitamos saber cuánto vale el máximo de N números distribuidos entre 900 y 950 y el segundo número más alto de N números distribuidos entre 900 y 1000. Aunque este cálculo es complicado, no es difícil convencerse de que, si N es suficientemente grande, es más probable que el precio final en la subasta de Vickrey sea mayor que en la de primer precio.

Por lo tanto, parece que Vickrey tiene todo a su favor: por un lado los participantes no tienen que estrujarse el cerebro para calcular su oferta ni tienen que adivinar o espiar las intenciones de sus contrincantes; en cuanto al subastador, tiene una alta probabilidad de conseguir un precio de venta más alto. Si es así, ¿por qué no se utiliza con más asiduidad? Una de las reticencias en contra de la subasta de segundo precio es que, si los participantes no son muy numerosos, el precio final de venta puede ser mucho menor que en la subasta de primer precio. Pero hay otra razón más sutil, que tiene que ver con la confianza de los participantes en la integridad del proceso. Si ganamos una subasta de Vickrey, pero no conocemos al resto de los participantes ni las ofertas que han enviado, ¿cómo sabemos que el segundo precio ha sido realmente el que nos dice el organizador?

Juan M.R. Parrondo

Incentivar la sinceridad

Alicia es presidente de una pequeña comunidad de cuatro vecinos: ella misma, Berta, Carlos y Daniel. Alicia pretende construir un ascensor en el edificio. El coste total asciende a 10.000 euros, pero no parece justo que todos paguen lo mismo. Algunos tienen una mayor necesidad del ascensor, como los que viven en los últimos pisos.

Alicia intenta en primer lugar un método bastante ingenuo. En una de las juntas pregunta públicamente a cada vecino (incluyéndose a sí misma en la encuesta) cuánto estaría dispuesto a pagar para tener el ascensor. Enseguida se da cuenta de que todos tienden a decir una cifra más baja de la que realmente estarían dispuestos a pagar, para ahorrarse algún dinero o por miedo a que los otros hagan lo mismo.

¿Cómo hacer que todos digan la verdad? Existe una ingeniosa solución a este problema, aunque no completamente satisfactoria. Se trata del llamado mecanismo de Vickrey-Clarke-Groves (VCG), inspirado en las subastas de Vickrey y diseñado por los economistas norteamericanos Edward Clarke y Theodore Groves. Lo abordamos aquí el mes pasado.

La idea es la siguiente. Alicia anuncia que va a preguntar a cada vecino lo que está dispuesto a pagar y que anotará las cantidades ofrecidas por cada uno, que llamaremos d_A, d_B, d_C y d_D. Si la suma de las cuatro ofertas alcanza los 10.000 euros, el ascensor se instalará. Hasta ahora el sistema es idéntico al primer intento ingenuo de Alicia. Lo que cambia en el mecanismo VCG es la contribución de cada vecino, que no pagará la cifra que ha declarado o ni siquiera una cantidad proporcional. Alicia pagará una cantidad dada por la siguiente fórmula:

$$c_A = 10.000 - (d_B + d_C + d_D).$$

Es decir, lo que paga Alicia no depende de lo que ella declara, sino sólo de lo que han ofrecido sus vecinos.

Esta extraña regla tiene una gran virtud: induce a decir la verdad. Supongamos que Alicia está dispuesta a pagar una cantidad v_A para tener el ascensor. Es decir, el valor real del ascensor para Alicia es exactamente v_A: si tuviera que pagar más estaría perdiendo dinero y si tuviera que pagar menos estaría ganando. (El mes pasado, al hablar de las subastas de Vickrey, comentamos con algo más de detalle el significado de este valor o "utilidad" de un bien, un concepto clave en economía.) Con más precisión, en caso de que el ascensor se instalara, Alicia "ganaría" una cantidad (que puede ser positiva o negativa):

$$U_A = v_A - c_A$$

¿Qué suma d_A debe ofrecer Alicia para maximizar su ganancia U_A? Ella no conoce el valor v_B, v_C, v_D, que tiene el ascensor para el resto de inquilinos, ni tampoco lo que éstos van a declarar. Pero sabe que el ascensor se instalará si el total de ofertas supera los 10.000 euros, es decir, si:

$$d_A > 10.000 - (d_B + d_C + d_D) = c_A.$$

Habrá, pues, ascensor si la oferta d_A de Alicia supera la cantidad c_A que ella misma pagará en caso de que la obra se acometa.

Si Alicia ofrece una suma d_A mayor que el valor real v_A, puede ocurrir que el resto de ofertas sea tal, que $d_A > c_A > v_A$. En este caso el ascensor se instalará y Alicia perderá dinero. Si, por el contrario, ofrece una suma d_A menor que el valor real v_A, puede ocurrir que el resto de ofertas sea tal que $d_A < c_A < v_A$, con lo que el ascensor no se construiría a pesar de que, si se hiciera, Alicia ganaría una cantidad U_A positiva, porque su contribución c_A sería inferior al valor v_A que el ascensor tiene para ella. Por lo tanto, la decisión óptima de Alicia es decir la verdad: declarar exactamente el valor real v_A. Lo mismo ocurre para el resto de vecinos.

El mecanismo VCG está inteligentemente diseñado para que mentir sea una mala estrategia. Si usted leyó esta sección el mes pasado, podrá encontrar una cierta similitud entre el mecanismo VCG y la subasta de Vickrey o de segundo precio. De hecho, hay también una subasta VCG que es una generalización de la subasta de segundo precio y que se basa en la misma idea que acabamos de ver: la puja sirve para decidir el adjudicatario, pero la cantidad a pagar no depende de su puja, sino de la del resto de los participantes en la subasta.

Sin embargo, en nuestro ejemplo del ascensor, como en cualquier problema de adquisición de bienes públicos, el mecanismo VCG adolece de un serio inconveniente. Bruno no paga la cantidad que ofreció, d_B, sino una cantidad c_B dada por una fórmula análoga a la de Alicia:

$$c_B = 10.000 - (d_A + d_C + d_D)$$

y así sucede con el resto de los vecinos. Por lo tanto, si se decide hacer la obra, la cantidad total que pagan nuestros cuatro amigos es:

$$C = c_A + c_B + c_C + c_D = 40.000 - 3D$$

en donde D es el total ofertado: $D = d_A + d_B + d_C + d_D$. C es mayor que 10.000 euros sólo si D es menor que 10.000. Pero precisamente la condición para que el

ascensor se construya es que D sea mayor que 10.000 Luego, aunque todos hayan dicho la verdad y la suma de lo declarado supere el costo total, ¡la comunidad no tendrá dinero suficiente para hacer la obra! Algún lector avispado puede haber pensado en la siguiente solución: Alicia da las reglas del mecanismo VCG pero, una vez conocido lo que está dispuesto a pagar cada vecino, las cambia y pide a cada uno lo declarado. Por supuesto, esto es bastante tramposo y la simple sospecha de tal comportamiento rompería el delicado balance sobre el que se basa el mecanismo VCG. Por ello, esta solución es inadmisible.

Una solución más honrada y práctica es completar lo que falta con algún fondo de reserva. Al fin y al cabo, en el caso de que la obra se apruebe, la suma D de lo ofertado no será muy superior a los 10.000 euros. Cada uno de los cuatro vecinos sabe que, si se repartiera el gasto en partes iguales, tendría que pagar 2500 euros. A esta cifra le añadirá o restará alguna cantidad pequeña, dependiendo de su necesidad de ascensor, con lo que la oferta total D estará cercana a los 10.000 y la suma C de las contribuciones será sólo ligeramente inferior al gasto total de 10.000 euros.

A pesar de sus limitaciones, el mecanismo VCG goza de un crédito considerable. Constituye uno de los pilares sobre los que se ha desarrollado todo un campo de la economía: la teoría de *mecanismos de revelación de la demanda*, que se puede aplicar en muy distintas situaciones como decisiones colectivas, subastas, el reparto de pistas en aeropuertos o incluso la regulación del tráfico de datos en Internet.

Las decisiones colectivas en donde se aplican estos mecanismos pueden ser muy variadas. Imaginemos un grupo de amigos que tienen que decidir qué película verán el próximo fin de semana. Supongamos que podemos cuantificar las preferencias de cada uno de ellos en términos monetarios y que éstas son las que se especifican en la siguiente tabla:

Casablanca		*El séptimo sello*	
Alicia	4 €	Daniel	2 €
Berta	12 €	Federico	5 €
		Esther	6 €
Carlos	9 €	Gabriel	7 €
Total	**25 €**	**Total**	**20 €**

Supondremos también que el beneficio de cada individuo es nulo si se va a ver la película que él no desea ver. Aunque la mayoría prefiere *El séptimo sello*, los que quieren *Casablanca* lo desean con más intensidad. Por lo tanto, si quiere maximizar el "beneficio total", el grupo debería ver *Casablanca*. El problema es de nuevo cómo obtener la tabla anterior. Si pedimos a cada miembro del grupo que cuantifique su deseo de ver la película, tenderán a exagerar, puesto que saben que la decisión final depende de la cantidad declarada.

Edward Clarke diseñó un mecanismo para obtener declaraciones sinceras, penalizando las exageraciones con un "impuesto": la *tasa de Clarke*. En primer lugar se calcula la diferencia ΔD entre los dos totales, que en nuestro ejemplo es de 5 euros. Los que están en el bando ganador y han declarado una cantidad superior a esta diferencia deben pagar la tasa, que es la diferencia entre su declaración y ΔD. En nuestro ejemplo, los únicos que tienen que pagar son Berta y Carlos, que abonarán, respectivamente, 7 y 4 euros. Con este "castigo", se puede comprobar que la estrategia óptima de todos los miembros del grupo es decir la verdad. Alicia no tiene interés ninguno en declarar más de los 4 euros con los que valora ir a ver *Casablanca*, puesto que su opción es ya ganadora. Si Berta aumenta su declaración y dice que para ella ver *Casablanca* equivale a un beneficio de 13 euros, acabará pagando la misma tasa. Y lo mismo ocurre si disminuye la cantidad declarada (siempre que no la baje tanto como para que su opción pierda). ¿Qué ocurre con los que están en el bando perdedor? Si, por ejemplo, Gabriel hubiera aumentado su declaración hasta 13 euros, el grupo acabaría asistiendo al pase de *El séptimo sello*, pero ΔD sería 1 euro y Gabriel tendría que pagar 12 euros, con lo que su beneficio neto sería negativo: $7 - 12 = -5$.

En realidad, éste es un caso particular del mecanismo VCG. Si d es mi declaración, D_m indica el total de mi opción y D_c el total de la opción contraria, la tasa de Clarke es:

$$T = d - D_m + D_c = D_c - D_m^*$$

en donde D_m^* representa el total de las declaraciones de mi opción *sin contar la mía*. Como en el mecanismo VCG, la tasa no depende de mi declaración. Si el valor real que tiene mi opción para mí es v, entonces mi beneficio será $U = v - T$ cuando gana mi opción y cero si pierde. Pero mi opción es ganadora si $D_m > D_c$, es decir, si $d > D_c - D_m^* = T$. La situación es idéntica al caso del ascensor: mi beneficio constituye el valor real menos la tasa si la declaración es mayor que la tasa y cero en otro caso.

Uno de los inconvenientes de este mecanismo de toma de decisiones radica en su vulnerabilidad ante coaliciones maliciosas. Federico y Esther podrían ponerse de acuerdo para declarar ambos 20 euros. El total para *El séptimo sello* sería entonces 49 y se convertiría en la opción ganadora. Pero ΔD sería ahora 24, con lo que nadie pagaría tasa alguna. Este aumento concertado de las declaraciones corre su riesgo, pero en general distorsiona el mecanismo de modo que la estrategia "sincera" deja de ser óptima. Por lo tanto, para que el mecanismo resulte eficaz, hay que evitar coaliciones mediante declaraciones secretas y sin posibilidad de negociación previa.

parr@seneca.fis.ucm.es

Juan M.R. Parrondo

El número *h*

Desde hace unos meses la comunidad científica está sobresaltada. En los congresos, convenciones o simples charlas de café entre investigadores y profesores de universidad, no se habla de otra cosa: el número *h*. ¿Es una nueva constante fundamental de la Naturaleza?, ¿la masa total del universo?, ¿alguna cantidad crucial en el origen de la vida? No. Se trata de algo bastante más mundano, pero capaz de despertar tantas pasiones entre los investigadores como cualquiera de los grandes enigmas de la ciencia.

El número *h* es un modo simple de evaluar la calidad de un científico. La idea original es de Jorge Hirsch, un físico argentino, profesor de la Universidad de California. Se publicó en septiembre del año pasado en los *Proceedings of the National Academy of Science*. Desde entonces, no hay científico que no haya calculado secretamente su número *h* y el de sus más queridos y odiados compañeros.

Siempre es útil disponer de una herramienta objetiva para evaluar la calidad de alguna actividad. Pero en el caso de la ciencia, una herramienta de este tipo se ha hecho indispensable por varias razones. En primer lugar, porque la mayor parte de los recursos destinados a la investigación básica son públicos y se necesitan criterios objetivos para distribuirlos. Administradores y evaluadores académicos tienen constantemente que repartir fondos, conceder becas y contratos o aprobar proyectos. En una empresa, la productividad de un departamento, incluido el de investigación y desarrollo, puede evaluarse con relativa facilidad en términos económicos. Sin embargo, lo que la investigación básica aporta a la sociedad no es siempre cuantificable económicamente o lo es a muy largo plazo.

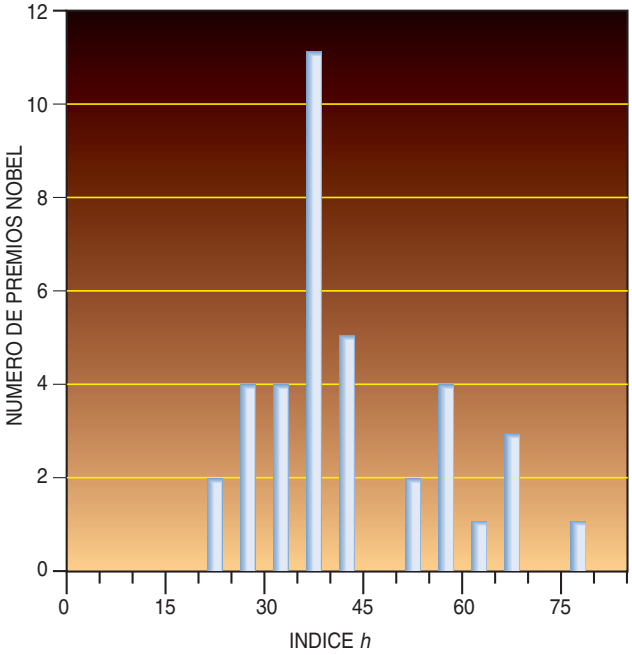
En segundo lugar, la creciente especialización hace cada vez más difícil este tipo de evaluaciones. En muchas ocasiones, la relevancia de un trabajo sólo puede ser juzgada con verdadera competencia por los contados expertos que investigan en ese mismo tema y que probablemente tendrán intereses compartidos o en conflicto con el autor de dicho trabajo, cuando no simples vínculos de amistad o enemistad.

Muchos consideran que estas razones reflejan la crisis de la ciencia contemporánea. ¿Cómo es posible que no sepamos decidir si una contribución científica es relevante o no? ¿Cómo es posible que sólo un grupo muy reducido de colegas sea capaz de semejante decisión? ¿Algo que *sólo* es relevante para unos pocos científicos que trabajan en el mismo tema no es, por definición, irrelevante para la ciencia y la sociedad? No hay respuesta sencilla.

La ciencia se ha convertido en una macroindustria en la que muchas pequeñas unidades trabajan a destajo en problemas insignificantes. Pero el conocimiento que se genera en estas modestas "unidades de producción" se transmite en ocasiones hacia grupos de científicos que se dedican a problemas cruciales o con más aplicaciones; por ejemplo, el origen del universo y el diagnóstico del cáncer, respectivamente. Es difícil prever cuáles de los miles de pequeños avances científicos repercutirán de un modo determinante en el futuro.

Una de las herramientas básicas para evaluar el impacto de un trabajo científico es el "número de citas", es decir, el número de artículos posteriores que lo mencionan. Cuando en un artículo citamos un trabajo anterior, significa que lo hemos utilizado, porque en él se desarrollaba una técnica que hemos empleado en el nuestro, porque se planteó un problema que ahora resolvemos o por cualquier otra razón. Así pues, las citas miden el impacto científico de un trabajo, sin que ello suponga, por necesidad, una calidad indiscutible. En ocasiones, un artículo reiteradamente citado ha resultado ser erróneo; cabe también que se le traiga a colación frecuente porque haya dejaba abiertas cuestiones fáciles de resolver. En cambio, puede muy bien acontecer que un artículo que cierre definitivamente un problema, aunque se trate de un problema de capital importancia, se cite menos que un artículo de calidad inferior.

A pesar de todo, el número de citas sigue siendo el parámetro común para medir la calidad de un trabajo. En 1960, Eugene Garfield, experto en bibliometría, creó el *Science Citation Index* (SCI), un índice o registro de



Histograma de los números *h* de los ganadores del premio Nobel de Física en los últimos 20 años.

las citas de los artículos científicos publicados en las principales revistas internacionales. El índice continúa siendo una herramienta fundamental en la investigación. Imagine el lector que ha encontrado un artículo de 1980 con unos resultados interesantes sobre los que quiere profundizar. El SCI le permite saber todos los artículos posteriores a 1980 que han citado ese trabajo. Con ello puede ponerse completamente al día sobre el asunto en cuestión.

Pero el interés del SCI de Garfield ha aumentado, asimismo, por su utilidad para evaluar la trayectoria completa de un investigador. ¿Cómo? Se han propuesto diversos indicadores, como el número total de citas o el número de citas por artículo. En su artículo, Hirsch aborda las limitaciones de estos indicadores y propone el suyo. Para el argentino, el número *h* de un científico es el máximo número del que puede decir: "tengo *h* artículos citado al menos *h* veces". Aunque su definición parezca extraña, es muy fácil encontrar el número *h* utilizando el SCI. Basta buscar todos los artículos de un autor y ordenarlos por número de citas. En la lista resultante, el artículo en donde el número de orden y el número de citas se cruzan nos proporciona el número *h* del autor. El procedimiento es tan sencillo, que en 10 minutos uno puede buscar el número *h* de todo su departamento (siempre que el nombre sea fácilmente reconocible: es bastante difícil encontrar el *h* de un Rodríguez o un Sánchez). En esa sencillez de "cálculo" se encierra el secreto del éxito de la idea de Hirsch.

Pero, ¿es el número *h* un buen indicador de la calidad de un científico? Hirsch ha calculado el número *h* de muchos físicos. Al parecer, en la cúspide se encuentra Edward Witten, físico matemático actualmente en Princeton, con un $h = 110$, es decir, tiene 110 artículos con al menos 110 citas. Más de 100 artículos de investigación constituyen una trayectoria científica bastante sólida. Un artículo con más de 100 citas se puede considerar un éxito profesional. ¡Las dos cosas a la vez, 110 artículos con más de 110 citas, es algo realmente impresionante! Sin embargo, Witten no ha conseguido el premio Nobel, al menos todavía, aunque sí la medalla Fields, máximo galardón en el ámbirto de la matemática.

De los siete físicos con mayor *h* hay cuatro premios Nobel: Anderson ($h = 91$), de Gennes ($h = 79$), Wilczek ($h = 68$) y Gross ($h = 66$). Completan el trío restante el mencionado Witten y dos físicos de materia condensada, Cohen ($h = 94$) y el español Manuel Cardona ($h = 86$), director del Instituto Max Planck del Estado Sólido en Stuttgart. En la figura podemos ver también un histograma de los números *h* de los premios Nobel de Física de los últimos 20 años, tomada del artículo de Hirsch. El índice *h* presenta una considerable dispersión, variando desde 22 hasta 79, pero está bastante concentrado en torno a 40.

Un físico con un número *h* en torno a 40 tiene, sin duda, una trayectoria influyente, en el sentido de que sus trabajos han dictado líneas de investigación seguidas por muchos otros.

Los *h* bajos son otra cuestión. Puede haber científicos con resultados muy relevantes y, sin embargo, con una trayectoria irregular. Recordemos a Mitchel Feigenbaum, quien encontró la primera ruta al caos conocida, aplicando además en el campo de los sistemas dinámicos una técnica desarrollada en la física estadística. Su contribución es muy relevante no sólo para la física, sino para todas las ciencias. Sin embargo, tiene un modesto $h = 16$, a pesar de que su principal trabajo sobre caos supera las 1700 citas.

En mi opinión, el índice *h* no es concluyente cuando se trata de evaluar o comprar trayectorias individuales, pero sí puede ser útil para análisis más generales. Por ejemplo, Hirsch deduce de sus propias observaciones que el *h* típico para conseguir una plaza de profesor permanente en una buena universidad de los Estados Unidos es 11, y 18 para una cátedra (siempre en el campo de la física). En las últimas pruebas para acceder al cuerpo de catedráticos en nuestro país, son esos los números que se han barajado. De hecho, en una de las disciplinas de mayor desarrollo, la física de la materia condensada, un candidato que no consiguió la plaza tenía un $h = 21$ (los que sí la consiguieron tenían números *h* parecidos). Al mismo tiempo, en nuestras universidades hay muchos catedráticos de física que no llegan a un $h = 5$. Esto no hace más que indicar el enorme progreso de la investigación científica española en las últimas décadas y la necesidad urgente de renovar el personal investigador o al menos de dar cabida a toda una generación de científicos de muy alto nivel que necesitan contratos y medios para desarrollar su trabajo.

El número *h* se puede calcular también para instituciones: será el máximo número del que se puede decir "hay *h* artículos de nuestra institución con al menos *h* citas". El número *h* de distintas universidades españolas y norteamericanas se muestra en esta tabla:

Universidad	*h*
MIT	470
Stanford	419
Berkeley	328
Harvard	307
Alabama	287
Maine	106
Autónoma de Madrid	152
Barcelona	141
Autónoma de Barcelona	98
Complutense de Madrid	96

De la tabla se desprende que *h* es un buen indicador del nivel de investigación de cada universidad. Algunas universidades españolas de primer rango podrían compararse a una universidad de nivel medio-bajo estadounidense, como la de Maine, a pesar de que nuestros científicos están, en media, a un nivel similar al de los científicos norteamericanos. Estos datos no hacen más que plasmar la necesidad de que los científicos jóvenes puedan desarrollar plenamente su carrera profesional en España.

JUEGOS MATEMÁTICOS

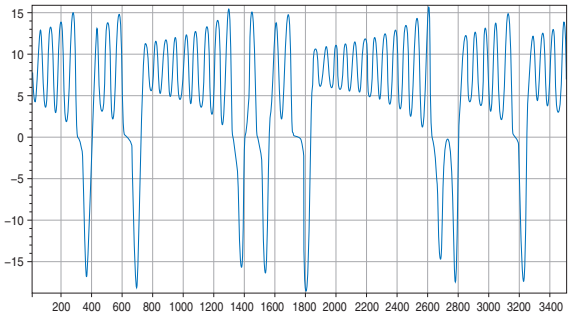
Juan M.R. Parrondo

Caos, recurrencia y consonancia musical

Desde que Isaac Newton creara el cálculo infinitesimal para estudiar el movimiento de los astros, una de las aplicaciones más fructíferas de la matemática ha sido el estudio del cambio. Sabemos cómo expresar matemáticamente las leyes que rigen la evolución de un sistema, mediante las llamadas *ecuaciones diferenciales*, y conocemos un sinfín de técnicas para resolverlas. También disponemos de una herramienta muy potente, la *transformada de Fourier*, que nos permite detectar periodicidades en el comportamiento de una magnitud.

Sin embargo, después de tres siglos de desarrollo continuado, y en gran parte gracias a la aparición de los ordenadores, el estudio de los sistemas que cambian en el tiempo experimentó una revolución con el descubrimiento de los *sistemas caóticos*. Fue el meteorólogo Edward N. Lorenz quien, al estudiar un modelo muy simple de la atmósfera, descubrió que una pequeña perturbación podía modificar de forma drástica el comportamiento futuro del sistema. En la figura 1 podemos ver cómo varía una de las variables del modelo de Lorenz. En períodos breves de tiempo la variable oscila de forma regular, pero en ocasiones se aleja de este comportamiento oscilatorio de un modo aparentemente imprevisible.

En torno a los sistemas caóticos surgieron nuevas técnicas para analizar cómo cambia una magnitud con el tiempo. Una técnica muy conocida, desarrollada por Jean Pierre Eckmann, de la Universidad de Ginebra, es el *diagrama de recurrencia*, que se aplica a series de datos. La idea es bastante simple. Si tenemos una serie de datos x_1, x_2, x_3,..., x_N, el diagrama de recurrencia consiste en una cuadrícula con $N \times N$ casillas en donde la luminosidad de la casilla (i,j) es inversamente proporcional a la distancia entre el punto x_i y el x_j.

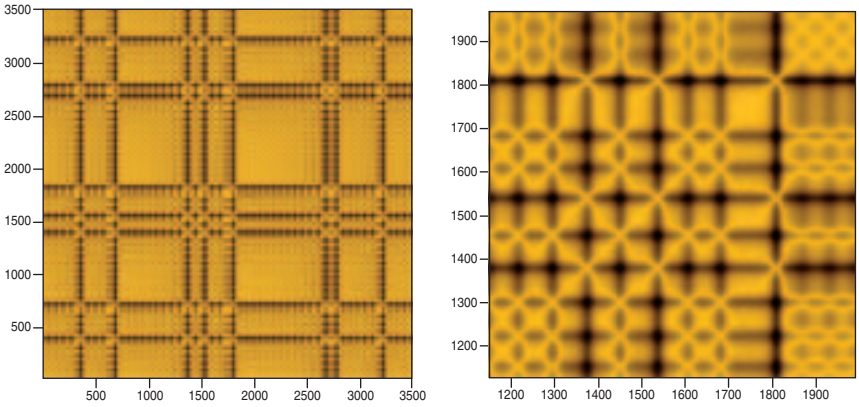


1. Evolución de una de las variables del modelo de Lorenz.

En la figura 2 podemos ver el diagrama de recurrencia de una de las variables del modelo de Lorenz y un detalle de dicho diagrama. La diagonal es siempre muy luminosa, puesto que muestra la distancia entre puntos que son iguales ($i = j$), es decir, a una distancia nula. Las zonas oscuras del diagrama corresponden a puntos muy distantes entre sí. La alternancia entre zonas claras y oscuras nos indica que la magnitud oscila. Por otra parte, en el diagrama no detectamos una periodicidad entre zonas claras y oscuras a escalas grandes de tiempo, debido al carácter caótico del sistema. En el detalle se pueden observar unos patrones que no son repetitivos, resultado también de esa mezcla de regularidad y complejidad característica de algunos sistemas caóticos.

El diagrama de recurrencia es en principio una herramienta meramente visual. Permite hacerse una idea de cómo evoluciona la variable en cuestión. En la figura 3 (*izquierda*) hemos representado el diagrama de recurrencia de una señal perfectamente periódica, sen(t). En este caso los patrones son también periódicos, como cabría esperar. Es interesante comprobar cómo se modifica el diagrama cuando añadimos ruido, es decir, una cierta cantidad aleatoria, a la señal periódica original. El resultado se muestra en la figura 3 (*derecha*). El diagrama se ha difuminado ligeramente y ha aparecido una textura en forma de malla, más visible en las zonas oscuras. Sin embargo, la estructura de la variable original se ha mantenido. Por lo tanto, el diagrama de recurrencia es útil para analizar variables aunque estén perturbadas por algún ruido.

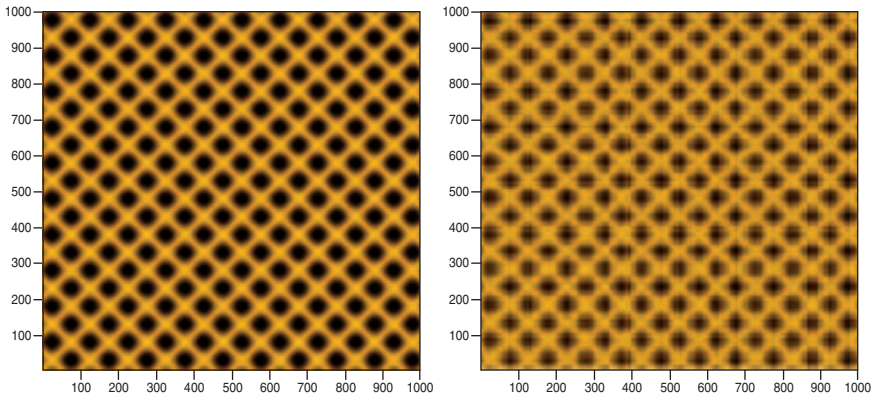
En la figura 4 mostramos diagramas de series de datos reales. El de la izquierda se ha obtenido con el número medio de manchas solares por año, desde 1700 hasta 1979 (280 datos). Se aprecia cierta periodicidad de unos 11 años, bien conocida. Pero, al igual que en el modelo de



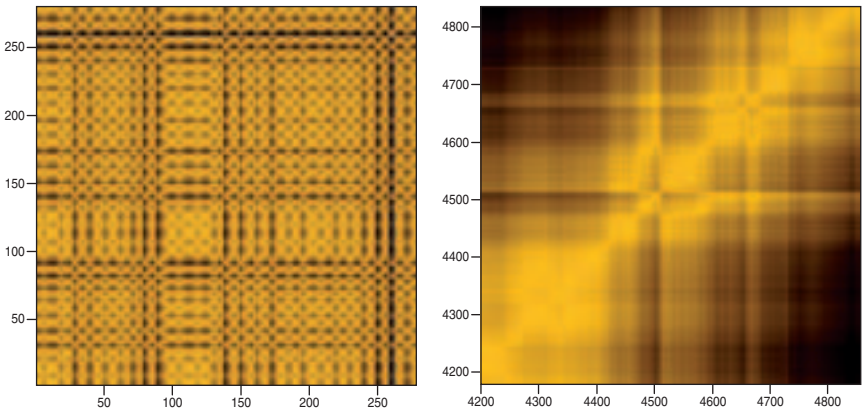
2. El diagrama de recurrencia de la variable del modelo de Lorenz que se representa en la figura 1. En la figura de la izquierda se puede ver el diagrama completo (3500 puntos) y en el de la derecha un detalle en torno a la diagonal (del punto 1150 al 2000). Los números de los ejes indican, como en las gráficas siguientes, el número de dato en la serie de datos.

Lorenz, esta periodicidad presenta irregularidades considerables, lo que nos indica que detrás de la evolución de las manchas se puede esconder una dinámica caótica. Finalmente, el diagrama de la derecha es el del índice bursátil Dow-Jones semanal, desde 1900 hasta 1996 (4977 datos). La textura es más uniforme, como corresponde a sistemas muy aleatorios. Vemos también que el diagrama se oscurece en la esquina superior izquierda y en la inferior derecha. Por una razón: los puntos alejados en el tiempo están también alejados entre sí, debido a que el Dow-Jones, a pesar de sus fluctuaciones, tiende a crecer en el tiempo.

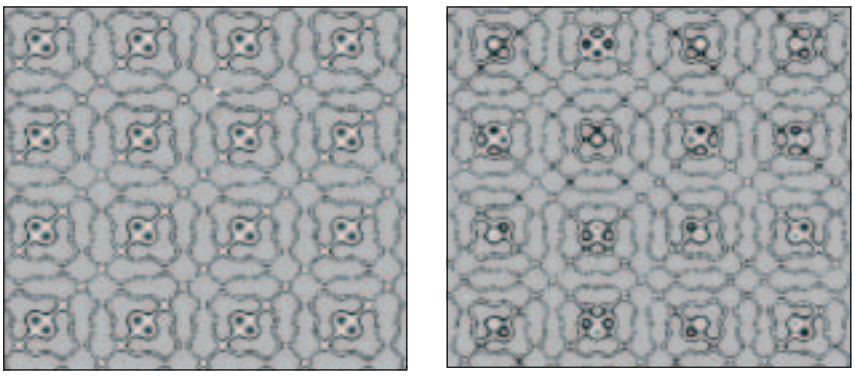
Los diagramas de recurrencia son una herramienta muy popular en el análisis de series temporales. Existen varios programas informáticos para elaborarlos. Los diagramas que hemos mostrado están hechos con el programa gratuito *Visual Recurrence Analysis (VRA)* de Eugene Kononov, que se puede descargar de Internet y es muy fácil de utilizar. Si tienen una serie de datos, cualquiera que sea su origen, les recomiendo que los analicen con el VRA. Es muy sencillo, y quizá les ayude a entender mejor el comportamiento de su serie (aunque lo realmente difícil es interpretar los diagramas).



3. Diagramas de recurrencia de una señal perfectamente periódica (*izquierda*) y de la misma señal perturbada por un ruido (*derecha*).



4. Diagramas de recurrencia de la intensidad de las manchas solares (*izquierda*) y del índice Dow-Jones (*derecha*).



5. El diagrama de recurrencia para un intervalo consonante (*izquierda*) y uno disonante (*derecha*).

Recientemente, Lluis Lligoña, del Centro de Investigación Puig Rodó de Gerona, junto con investigadores italianos y norteamericanos, ha aplicado esta técnica a un viejo problema del que ya hemos hablado en esta sección: la consonancia musical (*La teoría matemática de la consonancia*, marzo de 2004). Para ello, ha generado un sonido mezcla de dos tonos puros: uno de frecuencia fija y otro que va ascendiendo de forma suave, un *glissando* que barre todos los intervalos, desde el unísono (misma nota) con el primero hasta la octava.

En la figura 5 se pueden ver algunos de los resultados de Lligoña, aparecidos en el *Electronic Journal of Theoretical Physics*. Son dos detalles del diagrama de recurrencia. El de la izquierda corresponde a una zona en donde el intervalo es consonante, 2000 puntos en torno a la quinta justa, intervalo en donde las frecuencias están en una relación 3:2, mientras que el de la derecha está centrado alrededor de la quinta disminuida, un intervalo disonante de frecuencias 64:45. En la figura se aprecia fácilmente que la disonancia musical produce también una "disonancia visual". Lligoña y sus colaboradores han estudiado de forma cuantitativa algunas características del diagrama, como la densidad de puntos claros, y las han relacionado con la consonancia de los intervalos. Se consigue así una explicación de la consonancia muy diferente de la aceptada hasta la fecha, debida a Plomp y Levelt (*véase Juegos* de marzo de 2004) y que no es aplicable a tonos puros. No obstante, la teoría de Plomp y Levelt ha explicado fenómenos más complicados, como la dependencia de la consonancia con el timbre. Quizá los diagramas de recurrencia, aplicados a sonidos con timbre en lugar de a tonos puros, arrojen nueva luz sobre esta cuestión.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

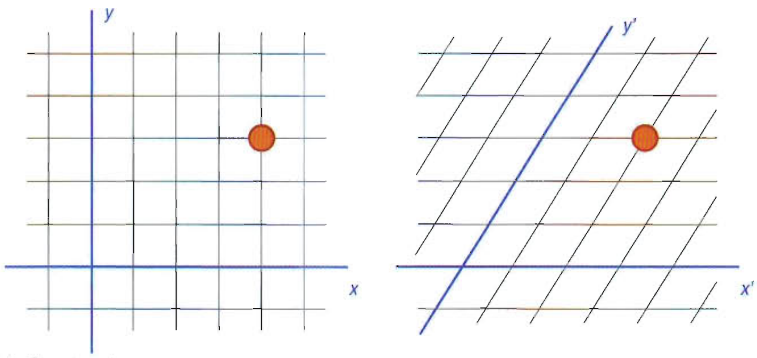
El espacio-tiempo

Cuando en física se dice que el espacio-tiempo tiene cuatro dimensiones, queremos decir simplemente que, para ubicar un suceso, basta con dar cuatro números: las tres coordenadas del punto espacial donde el suceso ocurre y el instante de tiempo en el que ocurre. La palabra "punto del espacio-tiempo" tiene un significado bastante ramplón: no es más que un lugar y una fecha.

¿Por qué entonces tanto misterio en torno al espacio-tiempo y sus cuatro dimensiones? Desplegar el tiempo sobre una dimensión significa dotarlo de una cierta "espacialidad" y poner en un mismo nivel los acontecimientos pasados, presentes y futuros. Esto tiene un indudable atractivo. Pero el espacio-tiempo ha adquirido una significación mucho más potente e inesperada gracias a la teoría de la relatividad, al demostrar que ambos tipos de "coordenadas", las espaciales y la temporal, no son independientes entre sí, sino que se entremezclan de forma sorprendente y contraria a la intuición.

Antes de adentrarnos en los misterios del espacio-tiempo, conviene que recordemos las diferentes formas de asignar coordenadas a los puntos de un plano. La forma más habitual es trazar dos ejes perpendiculares, como en la figura 1 (*izquierda*). A partir de estos dos ejes se puede dibujar una malla formada por líneas paralelas a los ejes y separadas entre sí por una unidad de longitud, por ejemplo un centímetro. Esta malla define las coordenadas de cada punto. Las del punto rojo de la figura izquierda son (4 cm, 3 cm). De la misma forma, se pueden encontrar las coordenadas de todos los puntos de la malla, y en general, utilizando números decimales, de cualquier punto del plano.

Existen otras formas de asignar coordenadas que no son tan habituales. En la figura 1 (*derecha*) se muestra un ejemplo en el que los ejes no son perpendiculares. Sin embargo, se puede también dibujar una malla mediante líneas paralelas a los mismos y asignar coordenadas de forma similar a nuestro primer ejemplo. En este caso, el punto rojo (que es el mismo que en el dibujo de la izquierda) tiene coordenadas (2 cm, 3 cm). El punto es el mismo en los dos dibujos; sin embargo, sus coordenadas son distintas porque hemos cambiado los ejes. Este cambio es parecido a lo que le ocurre a una magnitud cuando se expresa en distintas unidades: 10 centímetros y 0,1 metros son la misma distancia, aunque su expresión numérica depende obviamente de las unidades. Con un punto del espacio pasa lo mismo: puede tener expresiones numéricas diferentes, es decir, coordenadas diferentes, según los ejes que utilicemos. Pero, ¿para qué necesitamos diferentes ejes? ¿No basta con los habituales ejes perpendiculares?

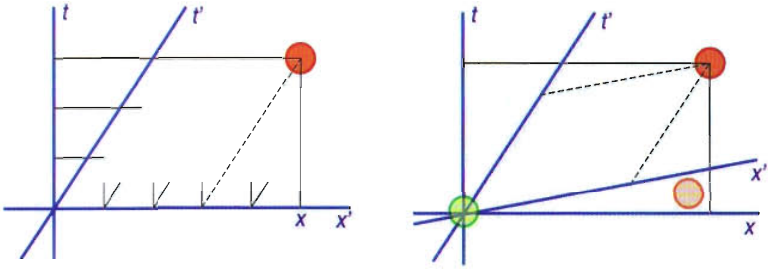


1. Coordenadas en el plano. El *mismo* punto (*círculo rojo*) tiene coordenadas $x = 4$, $y = 3$, con respecto a los ejes perpendiculares de la izquierda, y coordenadas $x' = 2$, $y' = 3$, con respecto a los ejes oblicuos de la derecha.

Utilizar distintos ejes sirve para entender cómo dos observadores pueden ver la misma realidad de distinta forma. Lo podemos comprobar con un ejemplo sencillo, en donde el espacio tiene sólo una dimensión y, por tanto, el espacio-tiempo es bidimensional. Supongamos que dos observadores sincronizan sus relojes en un determinado instante y lugar, es decir, en un determinado punto del espacio-tiempo. Los dos observadores colocan su origen de posiciones y de tiempos en este punto, que tendrá coordenadas (0,0) para ambos. Supongamos ahora que el primero de los observadores, O, permanece quieto y el segundo, O', se mueve a una velocidad v. Si el primero de ellos asigna unas coordenadas (x,t) a un punto del espacio-tiempo, ¿cuáles son las coordenadas (x',t') que asignará el segundo a ese mismo punto? La respuesta más intuitiva coincide con el análisis de Galileo. En primer lugar, el tiempo es universal: los dos observadores datarán un evento con la misma coordenada temporal. En otras palabras, hoy es 5 de mayo tanto para usted, que está sentado en su casa leyendo, como para mí, aunque esté viajando en un tren a 100 kilómetros por hora. Matemáticamente: $t' = t$. A la coordenada espacial no le ocurre lo mismo. Puesto que O' se mueve a velocidad v, verá alejarse el origen, que ya no tendrá coordenada espacial cero. La coordenada espacial para O' es $x' = x - vt$. En efecto, si la velocidad v es positiva, el observador O' se moverá hacia la derecha y verá cómo el origen $x = 0$ se aleja de él hacia la izquierda, de modo que le asignará una coordenada espacial negativa $x' = -vt$. La *transformación de Galileo*:

$$x' = x - vt \qquad t' = t$$

indica cómo se relacionan las coordenadas que el observador O y el observador O' asignan al *mismo* punto del espacio-tiempo. ¿Qué aspecto



2. Ejes de coordenadas en el espacio-tiempo para un observador en reposo y otro que se mueve hacia la derecha según Galileo (*izquierda*) y según Einstein (*derecha*).

tiene la transformación de Galileo? ¿Cómo son los ejes coordenados de *O* y de *O*' en el espacio-tiempo? La respuesta se muestra en la figura 2 (*izquierda*), para un observador *O*' que se mueve con respecto a *O* a una velocidad de 2/3 metros por segundo hacia la derecha. Como se puede apreciar, los ejes espaciales de ambos observadores coinciden, mientras que el eje temporal de *O*' está inclinado con respecto al de *O*. No he dibujado el mallado completo, como en la figura 1, sino sólo las líneas cercanas a los ejes, para no complicar demasiado la figura. El punto rojo tiene coordenadas $x = 5$, $t = 3$, para *O*, y $x' = 3$, $t' = 3$, para *O*', de acuerdo con las transformaciones de Galileo.

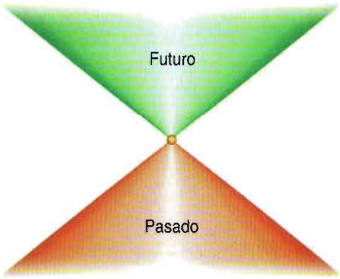
Observen que los segundos del eje temporal *t*' tienen una mayor longitud que los de *t*. Así debe ser si queremos reproducir gráficamente las transformaciones de Galileo. A primera vista es también curioso que los ejes espaciales de *O* y de *O*' coincidan y los ejes temporales difieran, mientras que ocurre lo contrario con las coordenadas. Reflexionando un poco, el lector puede comprobar que los ejes espaciales tienen que coincidir para que $t = t'$. De hecho, el eje temporal de *O*' está definido por los puntos con $x' = 0$, es decir, por la propia trayectoria de *O*'.

¿Qué ocurre en la teoría de la relatividad? Einstein descubrió que las transformaciones de Galileo debían ser sustituidas por las llamadas *transformaciones de Lorentz*:

$$x' = \frac{x - vt}{\sqrt{1-(v/c)^2}} \qquad t' = \frac{t - vx/c^2}{\sqrt{1-(v/c)^2}}$$

en donde *c* es la velocidad de la luz. Si la velocidad del observador *O*', *v*, es muy pequeña comparada con la velocidad de la luz, es fácil ver que las transformaciones de Lorentz son aproximadamente iguales a las de Galileo. Sin embargo, hay dos grandes diferencias entre ambas. En primer lugar, el denominador en las transformaciones de Lorentz hace que tanto el espacio como el tiempo se "contraigan" para *O*': las distancias serán más cortas y los objetos más pequeños en la dirección de su movimiento. La segunda y más importante diferencia es que el tiempo ya no es universal, sino que la coordenada temporal se mezcla con la espacial a través del término vx/c^2. En consecuencia, los ejes espaciales de los dos observadores ya no coinciden, como se ve en la gráfica de la derecha de la figura 2. El tiempo deja de ser absoluto. Por ejemplo, para *O*' los puntos que están sobre el eje *x*' son simultáneos, es decir, a todos ellos *O*' les asigna la misma coordenada temporal $t' = 0$. Sin embargo, para *O* esos mismos puntos tienen distinta coordenada temporal, creciente según nos movemos a lo largo del eje hacia la derecha. En pocas palabras, la simultaneidad es relativa. Incluso el orden temporal es relativo: para *O* el círculo rojo claro de la figura es posterior al verde, mientras que para *O*' es anterior en el tiempo.

Esta relatividad del orden temporal plantea enseguida un problema adicional: ¿qué ocurre si en un punto A del espacio-tiempo ocurre algo que es causa de un evento en un punto B? ¿Vería entonces *O*' el efecto antes de la causa? Uno de los principios básicos de la teoría de la relatividad impide este tipo de paradoja: nada puede viajar a una velocidad mayor que la de la luz, ni cuerpos materiales ni información. Por lo tanto, lo que ocurre en A puede ser causa de lo que ocurre en B sólo si estos dos puntos del espacio tiempo pueden "comunicarse" entre sí a través de un rayo de luz o de algo que viaje más despacio que la luz. Esta limitación se representa gráficamente en la figura 3. El punto amarillo en el centro del dibujo sólo puede influir sobre el triángulo verde en el espacio-tiempo y sólo puede ser influido por sucesos ocurridos en el triángulo rojo. Son en realidad, y respectivamente, el futuro y el pasado del punto amarillo, si entendemos por pasado aquello que podemos recordar o que nos puede influir y por futuro aquello en lo que aún tenemos la posibilidad de intervenir. Estos triángulos de causalidad no dependen del observador. Por lo tanto, entendidos de esta forma, el pasado y futuro de cualquier punto del espacio-tiempo son absolutos. Por ejemplo, el orden temporal del punto rojo claro y del verde en la figura 2 (*derecha*) es diferente para *O* y para *O*', como ya hemos visto. Sin embargo, el punto rojo no es futuro ni pasado del verde, y viceversa.



3. Pasado y futuro según la teoría de la relatividad.

Con ello la teoría de la relatividad recupera toda su coherencia. Sin embargo, hay algo en el espacio-tiempo einsteiniano que sigue siendo enigmático, especialmente cuando uno considera sucesos aleatorios. Nos parece que la tirada de un dado establece de forma nítida un "antes" y un "después", un universo antes de la tirada y un universo después de la tirada, incluso aunque puntos de ese universo no puedan verse en absoluto influidos por el resultado de la tirada. Sin embargo, un observador puede considerar la tirada anterior a cierto evento, mientras que otro observador la puede considerar posterior. La inclusión del azar en el espacio-tiempo einsteiniano parece problemática, a pesar de no producir paradojas de causalidad. A este tema dedicaremos probablemente alguna reflexión en el futuro.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Espaciotiempo y azar

El mes pasado discutimos la sorprendente estructura del espaciotiempo según la teoría especial de la relatividad. Tiempo y espacio se mezclan cuando tratamos de relacionar cómo ven el mundo observadores que viajan a distinta velocidad. Supongamos dos observadores, *O* en reposo y *O'* moviéndose a una velocidad *v*, que colocan su origen de posiciones y de tiempos en un mismo punto. Si *O* asigna unas coordenadas (x,t) a un punto del espaciotiempo, las coordenadas (x',t') que asignará *O'* son obviamente distintas. En la vieja mecánica de Galileo y Newton, la relación entre ambas coordenadas es:

$$x' = x - vt \qquad t' = t$$

mientras que, en la teoría de la relatividad especial, las coordenadas de los dos observadores se relacionan a través de las llamadas *transformaciones de Lorentz*:

$$x' = \frac{x - vt}{\sqrt{1-(v/c)^2}} \qquad t' = \frac{t - vx/c^2}{\sqrt{1-(v/c)^2}}$$

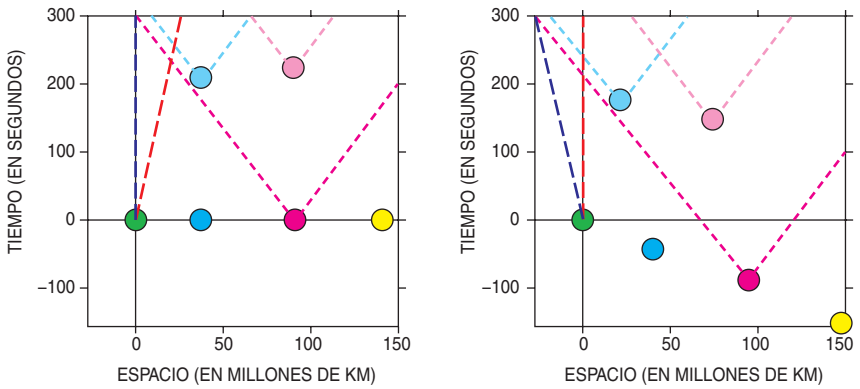
en donde *c* es la velocidad de la luz.

Vamos a analizar algunas de las propiedades sorprendentes de las transformaciones de Lorentz. En los dos diagramas de la figura 1 representamos varios puntos del espaciotiempo, tal y como los ven *O* (*izquierda*), en reposo, y *O'* (*derecha*), que viaja hacia la derecha a una velocidad de 90.000 kilómetros por segundo (0,3 veces la velocidad de la luz). He escogido las distancias de modo que los eventos estén localizados en la Tierra (*punto verde*), Venus (*puntos azules*), Mercurio (*puntos rojos*) y el Sol (*punto amarillo*). Recordemos que las dos figuras representan los *mismos* puntos, pero localizados y datados por *O* y por *O'*. He dibujado también las trayectorias de *O* (*en azul*) y de *O'* (*en rojo*) y los llamados *conos de luz* de algunos eventos, en concreto de los dos eventos en Mercurio y del segundo en Venus. Los conos son las dos líneas que parten de cada punto del espaciotiempo y representan la trayectoria de dos rayos de luz emitidos desde dicho punto. Estos conos tienen una propiedad fundamental: su forma no varía de un observador a otro. Siempre presentan el mismo ángulo debido a que en las transformaciones de Lorentz la velocidad de la luz es la misma para todos los observadores. Es una propiedad fácil de demostrar para el caso particular del cono de luz $x = ct$.

Observemos la serie de eventos que se encuentran a tiempo 0 para *O*, es decir, los cuatro puntos que están en el eje horizontal de la figura 1 (*izquierda*). Estos puntos son simultáneos para *O*, ya que les asigna la misma coordenada temporal, $t = 0$. Por ello, el eje horizontal se denomina la línea "ahora" para *O* en el momento en que se encuentra sobre el punto verde de la figura. Sin embargo, estos cuatro puntos no son simultáneos para *O'*, quien, por ejemplo, data el evento amarillo unos dos minutos y medio antes que el evento verde. La línea "ahora" de *O* es para *O'* una línea inclinada hacia abajo. A su vez, la línea "ahora" de *O'* es horizontal en la figura de la derecha, pero sería oblicua, inclinada hacia arriba, en la figura de la izquierda, es decir, en las coordenadas de *O*. Por supuesto, todas estas líneas "ahora" suben paulatinamente, a medida que el tiempo avanza para cada uno de los observadores, aunque conservando su inclinación.

La simultaneidad es relativa. Primer hallazgo de Einstein. Pero incluso el orden temporal es relativo. El evento azul superior en Venus ocurre para *O* en un tiempo $t = 100$, unos veinte segundos antes que el evento naranja que ocurre en Mercurio. Sin embargo, para *O'*, el segundo de estos dos eventos, es decir, el de Mercurio, es unos veinte segundos anterior al de Venus. Como vimos el mes pasado, esta relatividad del orden temporal de eventos no da lugar a paradojas causales. Si hay ambigüedad en el orden temporal de dos eventos, uno no puede influir sobre el otro. La ambigüedad sólo ocurre entre puntos del espaciotiempo que no se pueden comunicar entre sí, o que podrían comunicarse sólo mediante señales más rápidas que la luz, lo cual es imposible en el marco de la relatividad especial.

Cuando se consideran eventos aleatorios, como la tirada de un dado, siguen sin darse paradojas causales, pero la relatividad del tiempo produce una sensación inquietante.



1. Varios eventos en el espaciotiempo tal y cómo los ven dos observadores con distintas velocidades. Las trayectorias de los dos observadores se muestran como líneas discontinuas en rojo y azul partiendo del origen. También se muestran los conos de luz de algunos de los eventos.

Hay una definición de pasado y futuro puramente causal: pasado es todo aquello que puede influirme y futuro es todo aquello sobre lo que puedo influir. Pero creo que el azar establece una diferencia más fuerte entre pasado y futuro. Supongamos que tiro un dado a las 12:00 del día 15 de junio y el resultado es un 6. Tenemos la sensación de que, con la tirada, ha habido algo parecido a una "creación". Antes de la tirada el resultado era completamente incierto, no existía. La diferencia entre antes y después de las 12:00 del 15 de junio no estriba sólo en los posibles efectos del resultado de la tirada. El resultado, 6, no estaba "escrito" en ningún sitio. Se "escribe", comienza a existir, cuando realizamos la tirada.

En el espaciotiempo newtoniano el azar tiene fácil cabida. El universo avanza como un todo a lo largo del tiempo. Podemos representar este avance tapando con una hoja de papel el diagrama de espaciotiempo y moviéndola hacia arriba. Los puntos tapados son puntos que aún no han ocurrido y los descubiertos son los que ya han ocurrido. Así fluye el tiempo newtoniano.

¿Se puede trasladar la noción "ha ocurrido" al espaciotiempo einsteiniano? Volviendo a la figura 1, supongamos que alguien lanza un dado en Mercurio en $t = 0$ (*punto inferior azul de la figura*) y envía el resultado mediante un rayo de luz hacia la Tierra. El observador *O* recibe la información en el punto en donde se cortan su trayectoria (*la línea azul punteada*) y el cono de luz que parte de Mercurio. Hace sus cálculos y concluye que la tirada *ocurrió* justo cuando él estaba en el punto verde. A pesar de que *O* recibe la información en $t =$ 300 segundos, sólo puede recibirla si esta información se ha producido con anterioridad. Sin embargo, *O'*, quien también recibe el resultado, realiza el mismo tipo de argumento concluyendo que la tirada ocurrió en $t' = -100$ segundos. Que las dataciones de los dos observadores difieran no tiene nada de raro a estas alturas. Al fin y al cabo, t y t' son maneras de datar el *mismo* evento por parte de cada observador.

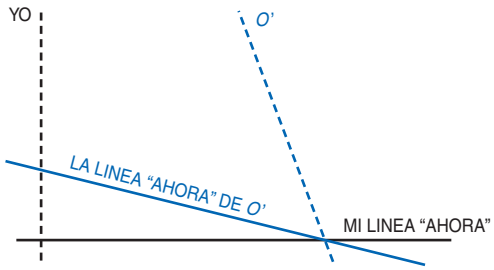
La cuestión se puede complicar un poco más. Supongamos dos tiradas de dado genuinamente aleatorias, que ocurren, respectivamente, en dos puntos A y B cuyo orden temporal es ambiguo. Para un observador *O*, la tirada A ocurre antes que la tirada B, mientras que para *O'* la B es anterior a la A. Es decir, *O* piensa que ha habido un universo en donde el resultado de A existía y el resultado de B no existía aún; mientras que *O'* pensará lo contrario. A pesar de que esta relatividad de lo que "ha ocurrido" no conduce a ninguna paradoja de causalidad, parece contradecir la sensación de que la historia del universo, en lo que se refiere a eventos aleatorios, "se escribe". En realidad este "se escribe" presupone ya la noción de que existe "un universo en un momento dado", lo cual contradice la teoría de la relatividad especial. No hay nada parecido a "el universo en tal o cual momento". Cualquier punto del espaciotiempo está en la línea "ahora" de algún observador suficientemente alejado.

En la figura 2 podemos ver esta situación de forma más explícita. La línea vertical es "mi historia", mi situación a lo largo del espacio tiempo. Muy alejado se encuentra otro observador, *O'*, que viaja hacia mí a una velocidad considerable. En la figura vemos mi línea "ahora" y la línea "ahora" de *O'*, que cruza mi historia, digamos, diez años más tarde que mi "ahora". Resulta entonces que, en mi "ahora", *O'* está pensando que los diez años siguientes de mi vida han ocurrido ya. La información de lo ocurrido en esos diez años le llegará mucho más tarde a *O'*, pero, cuando le llegue, concluirá que esos diez años ya habían ocurrido cuando él estaba en mi "ahora". Es decir, mis diez años siguientes están en cierto modo presentes en mi "ahora" a través de *O'*.

A diferencia del newtoniano, el espaciotiempo relativista parece implicar que "todo está escrito". Como decía Popper, el espaciotiempo einsteiniano se parece bastante a la cosmovisión de Parménides, para quien el tiempo era una ilusión. La cosmovisión opuesta, el "todo fluye" de Heráclito, que podríamos interpretar como "el futuro está por escribir", tiene una dudosa cabida en la relatividad especial, puesto que ni siquiera está claro qué entendemos por "futuro".

El filósofo Nuel Belnap desarrolló en un conocido artículo de 1992 una interesante teoría formal de ramificaciones en el espaciotiempo, pero sin aludir a ninguna clase de "escritura" paulatina de la historia del universo. En su teoría, los resultados de eventos aleatorios están ya "escritos" sobre el espaciotiempo ramificado.

Curiosamente, la cosmología relativista sí permite incluir la noción de un tiempo que fluye. Uno de sus puntos de partida, el llamado *principio cosmológico*, establece la existencia de ciertos observadores privilegiados, los llamados *observadores fundamentales*. Son observadores que se mueven igual que lo



2. En mi "ahora", *O'* piensa que los próximos diez años mi vida "ya han ocurrido".

hacen, en media, las galaxias que están a su alrededor y definen una línea de tiempo común a todos ellos, el *tiempo cósmico*, y también una espacio "ahora" común. Por ejemplo, todos los puntos del espaciotiempo en los que la temperatura de la radiación de fondo es la misma definirían un espacio de simultaneidad universal. Sería entonces posible considerar que la historia del universo se está escribiendo, al igual que en el espaciotiempo newtoniano, como una sucesión de estas simultaneidades y existirían observadores (que no serían fundamentales) para los cuales su línea de "ahora" tendría puntos que aún no se han escrito, que aún no han ocurrido.

En cualquier caso, creo que la noción "ha ocurrido" es puramente filosófica. Quiero decir con ello que lo más probable es que no tenga consecuencias observables. Aun así, profundizar en las relaciones entre el espaciotiempo y el azar sí puede ser útil para una de las tareas pendientes de la física actual: conciliar la relatividad especial y general con la mecánica cuántica, en la que el azar desempeña un papel fundamental.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Otras formas de contar

"El reciente censo invernal de aves acuáticas ha contabilizado en Doñana 164.131 ejemplares". "El número de ballenas jorobadas en Baja California es aproximadamente 1450". ¿No se han preguntado nunca cómo se obtienen estos números? ¿Hay alguien que cuenta pacientemente todas las aves o ballenas que pasan delante de sus narices? Y si es así, ¿cómo se las apaña para no contar dos veces el mismo animal?

La primera operación matemática que aprendemos es la de contar y cualquiera sabe hacerlo hoy en día. Sin embargo, lo que no todo el mundo sabe es que hay maneras más inteligentes de contar que la mera enumeración de objetos o individuos.

Para realizar censos de animales, se conoce desde el siglo XIX una técnica bastante ingeniosa: *el método de captura-recaptura*. Supongamos que queremos saber el número de peces de un estanque. Pescamos un cierto número de ellos, digamos 100 (cuanto más alto, mejor). Marcamos estos peces capturados de alguna forma y los soltamos de nuevo en el estanque. Esperamos un tiempo suficientemente largo para que los peces marcados se mezclen con el resto —por ejemplo, una semana—, y volvemos a capturar 100 peces. De ellos, algunos estarán marcados, por ser peces "recapturados", y otros no. La proporción de peces recapturados nos indicará la población total en el estanque. En efecto, si hay en total N peces y he marcado 100, la probabilidad de pescar en la segunda captura un pez marcado será $100/N$. Por lo tanto, el número medio o "esperado" de peces recapturados es $100 \times 100/N$. Si n es el número de peces que realmente hemos recapturado, una buena estimación de la población total es:

$$N = \frac{100 \times 100}{n}$$

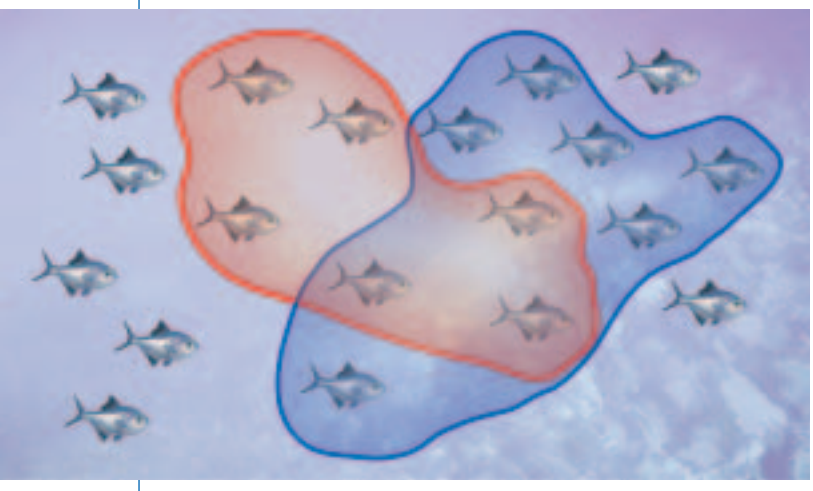
Es decir, si recapturo 20 peces, lo más probable es que la población total sea de unos 500. En general, si en la primera captura pesco n_1 peces y en la segunda n_2, de los cuales n_{12} resultan estar marcados, entonces la mejor estimación de la población total de peces es:

$$N = \frac{n_1 \times n_2}{n_{12}}$$

En la figura 1 podemos ver una representación esquemática del proceso.

Esta es la versión más simple del método de captura-recaptura. Tal y como lo hemos formulado, son necesarias varias condiciones para que la estimación sea válida. En primer lugar, no sólo el número de peces debe permanecer constante entre las dos capturas (si no fuera así, no tendría sentido calcular dicho número), sino que no puede haber muertes o nacimientos de peces. Si los hubiera, la fracción de peces marcados en la segunda captura sería menor que esa misma fracción en el momento de soltarlos al estanque, es decir, inmediatamente después de la primera captura, puesto que algunos habrán podido morir entre las dos capturas. Ocurriría lo mismo si los peces pudieran abandonar el lugar de estudio, por ejemplo si fuera un río en lugar de un estanque. La segunda suposición es que las capturas tienen que ser completamente aleatorias. Esta aleatoriedad es en ocasiones difícil de conseguir, especialmente en la segunda captura, porque los peces marcados pueden haber "aprendido" a evitar las redes. En general, hay que asegurarse de que los animales marcados no tienen mayor o menor tendencia a ser recapturados que sus congéneres no marcados.

Los métodos más avanzados de captura-recaptura son bastante más complejos de lo que hemos descrito aquí y tienen en cuenta la posibilidad de muertes y nacimientos o de migraciones. Se basan no sólo en dos capturas, sino en un número arbitrario de ellas, después de las cuales los animales vuelven a soltarse, que conforman una "historia" para cada animal capturado. Estas historias suelen denotarse mediante cadenas de ceros y unos: un uno significa que el animal ha sido capturado y un cero significa que no lo ha sido. Por ejemplo, si se han realizado cinco capturas, la historia 01100 indica que el animal ha sido capturado en la segunda y tercera de ellas y se ha "escapado" en el resto, es decir, en la primera, cuarta y quinta. Cada animal tiene una historia. Incluso se puede asociar la historia 00000 a los animales que



En la primera captura (*línea roja*) hemos obtenido 6 peces. En la segunda (*línea azul*) hemos obtenido 9, de los cuales hay 3 marcados. La estimación del número total de peces es $6 \times 9/3 = 18$, ligeramente inferior al valor real, 19.

no han sido cazados en ninguna de las cinco capturas, aunque, a diferencia del resto de historias, no sabemos el número de animales con historia 00000. De hecho, este número es normalmente la incógnita que queremos conocer y se suele estimar a partir del número de animales con historias que contienen al menos un uno, dato que sí conocemos perfectamente de nuestro registro de capturas.

El caso más sencillo es el de dos capturas, sin considerar muertes o migraciones, situación a la que se denomina *sistema cerrado*. Las cuatro posibles historias se muestran en la siguiente tabla:

HISTORIA	NUMERO DE ANIMALES	PROBABILIDAD	NUMERO ESPERADO DE ANIMALES
11	x_{11}	$p \times p$	$N \times p \times p$
01	x_{01}	$(1-p) \times p$	$N \times (1-p) \times p$
10	x_{10}	$p \times (1-p)$	$N \times p \times (1-p)$
00	x_{00}	$(1-p) \times (1-p)$	$N \times (1-p) \times (1-p)$

En la tabla se indican el número de animales con cada historia y la probabilidad de que un animal haya sido capturado de acuerdo con dicha historia. En el modelo más simple, en el que no hay muertes, nacimientos ni movimientos de población, la probabilidad de ser capturado en cada ocasión es $p = n/N$, siendo N el número total de animales y n el número de animales capturados, que aquí hemos supuesto, por simplicidad, el mismo para todas las capturas. La probabilidad de cada historia se obtiene multiplicando la probabilidad de ser capturado o no en cada una de las capturas. Finalmente, la cuarta columna es simplemente N por la probabilidad y nos indica el valor esperado de animales en cada historia. A partir de los valores esperados y de los valores reales, puede estimarse el parámetro p y la población total N utilizando métodos estadísticos, como el llamado de *máxima verosimilitud*. Cuando hay sólo dos capturas, el resultado es el que hemos obtenido en el ejemplo de los peces.

El problema se complica si consideramos tres capturas y aún más si el ecosistema que estamos estudiando es *abierto*, es decir, si admitimos la posibilidad de muertes o migraciones. En este último caso, se suele suponer que el número total de animales se mantiene constante, al menos de forma aproximada, gracias a nacimientos o inmigraciones que compensan las pérdidas. Sin embargo, estas compensaciones, como ya hemos mencionado, hacen que las probabilidades de cada historia cambien. En la tabla de la derecha se pueden ver estas probabilidades. En ellas aparece un nuevo parámetro, v, que es la probabilidad de seguir vivo entre captura y captura.

Aunque estas probabilidades son bastante más complejas que las anteriores, el lector con algunos conocimientos de probabilidad podrá derivar cada una de las entradas de la tabla sin excesiva dificultad. Al igual que en el caso anterior, varios métodos estadísticos permiten encontrar los parámetros p, v o ambos que mejor ajusten los datos de la tabla. Con ello se obtiene una información valiosa acerca del número de animales y de su probabilidad de supervivencia en un período dado. Hay modelos aún más complicados que éste y que permiten obtener otras características, como el número de animales en celo, su duración, los periodos de anidación, etc. Incluso circulan en Internet varios programas gratuitos, como SURGE o MARK, que realizan estas estimaciones de modo automático, si previamente se elige el modelo a utilizar.

La investigación matemática sobre el método de captura-recaptura continúa. Por ejemplo, en un trabajo reciente, Ricardo García-Pelayo, de la Universidad Politécnica de Madrid, proporciona un método para dar no sólo una buena estimación de la población total, sino también el intervalo de error del resultado. De hecho, gracias a este artículo es como he tenido conocimiento de este tema.

El método de captura-recaptura se ha aplicado también a los seres humanos. Uno de los pioneros de este método fue Laplace, quien lo utilizó para estimar la población de Francia a principios del XIX. Pero, en el caso de los humanos, ¿cómo se realizan las "capturas"? Basta cualquier listado de una muestra de los individuos que se quieren "contar", siempre que la muestra sea suficientemente aleatoria. Por ejemplo, Antonia Domingo Salvany y su grupo, de la Universidad Autónoma de Barcelona, han estimado el número de heroinómanos en Barcelona utilizando estos métodos. Las tres "capturas" consistieron en las listas de ingresos en urgencias, de solicitud de tratamiento de desintoxicación y de muerte por sobredosis, respectivamente. Según apareciera o no en cada una de estas tres listas, los investigadores asociaron a cada individuo una historia, tal y como hemos hecho en los ejemplos anteriores. Con estos datos, han obtenido finalmente la población total de toxicómanos, estuvieran o no registrados en alguna de las tres listas. En epidemiología la captura-recaptura es también un método muy utilizado para elaborar censos de distintas enfermedades.

HISTORIA	PROBABILIDAD (ECOSISTEMA CERRADO)	PROBABILIDAD (ECOSISTEMA ABIERTO)
111	$p \times p \times p$	$p \times vp \times vp$
110	$p \times p \times (1-p)$	$p \times vp \times [v(1-p) + (1-v)]$
101	$p \times (1-p) \times p$	$p \times v(1-p) \times vp$
100	$p \times (1-p) \times (1-p)$	$p \times [v(1-p) \times v(1-p) + v(1-p)(1-v)+(1-v)]$
011	$(1-p) \times p \times p$	$(1-p) \times vp \times vp$
010	$(1-p) \times p \times (1-p)$	$(1-p) \times vp \times [v(1-p) + (1-v)]$
001	$(1-p) \times (1-p) \times p$	$(1-p) \times v(1-p) \times vp$
000	$(1-p) \times (1-p) \times (1-p)$	$(1-p) \times [v(1-p) \times v(1-p) + v(1-p)(1-v)+(1-v)]$

En cualquier caso, el método está también sujeto a ciertas críticas. Las hipótesis en que se basa son en ocasiones demasiado restrictivas y, aun cumpliéndose, el error de la estimación final puede ser excesivo. Por otro lado, hay especies y ecosistemas en los que resultan mejores otras "formas de contar". Por ejemplo, cuando hay pocos individuos y se encuentran dispersos, es mejor realizar el recuento por pura inspección en las distintas zonas en donde habitan los animales. La Sociedad Española de Ornitología organiza cada año este tipo de censos para ciertas especies de aves en peligro de extinción con la colaboración de voluntarios de todo el país.

Juan M.R. Parrondo

Ganancia segura

Como supuesto experto en probabilidad y juegos de azar, una de las preguntas que más veces me han hecho en los últimos años es si conozco algún método para ganar en un casino o en la bolsa. La respuesta, invariablemente, es negativa. Sin embargo, esta vez tengo buenas noticias. Voy a revelarles por fin un método con el que, en principio, se puede ganar con absoluta certeza en ciertas apuestas deportivas que ahora abundan en Internet. No hay trampa ni cartón. Pero no se hagan excesivas ilusiones: el método requiere ciertas condiciones que son muy difíciles de cumplir, debido a las comisiones de las casas de apuestas.

La idea consiste en aprovechar las diferencias en el pago que dos casas de apuestas distintas ofrecen por el mismo resultado de un acontecimiento deportivo. En efecto, estos pagos no son siempre iguales en todas las casas e incluso pueden cambiar con el tiempo en una misma casa, puesto que dependen de la cantidad apostada a cada resultado por *todos* los participantes en la apuesta. Los pagos se suelen indicar con un número mayor que uno, la *cuota*, que es el dinero que se recibe por cada euro apostado. Cada resultado tiene su propia cuota, que será en general más alta cuanto más improbable sea.

El método de ganancia segura es simplísimo a primera vista. Supongamos que, rastreando por Internet, encontramos una eliminatoria de un campeonato de fútbol para la que dos casas de apuestas ofrecen las siguientes cuotas:

	Para A	Para B
Casa de apuestas 1	2	2
Casa de apuestas 2	2,22	1,82

Podríamos entonces apostar 1,10 euros a B en la casa número 1, y 1 euro a A en la número 2. La apuesta total sería de 2,10 euros. Si el resultado fuera A, recibiríamos $1 \times 2{,}22 = 2{,}22$, ganando 12 céntimos de euro. Por otro lado, si el resultado fuera B, entonces recibiríamos $1{,}10 \times 2 = 2{,}20$, ganando 10 céntimos. Es decir, ganamos sea cual sea el resultado.

Pero quizá hayamos pecado de optimistas. ¿Es posible encontrar en Internet las cuotas de la tabla anterior u otras que permitan obtener un beneficio seguro? Para responder a esta pregunta tenemos que conocer el modo en que las casas de apuestas establecen sus cuotas.

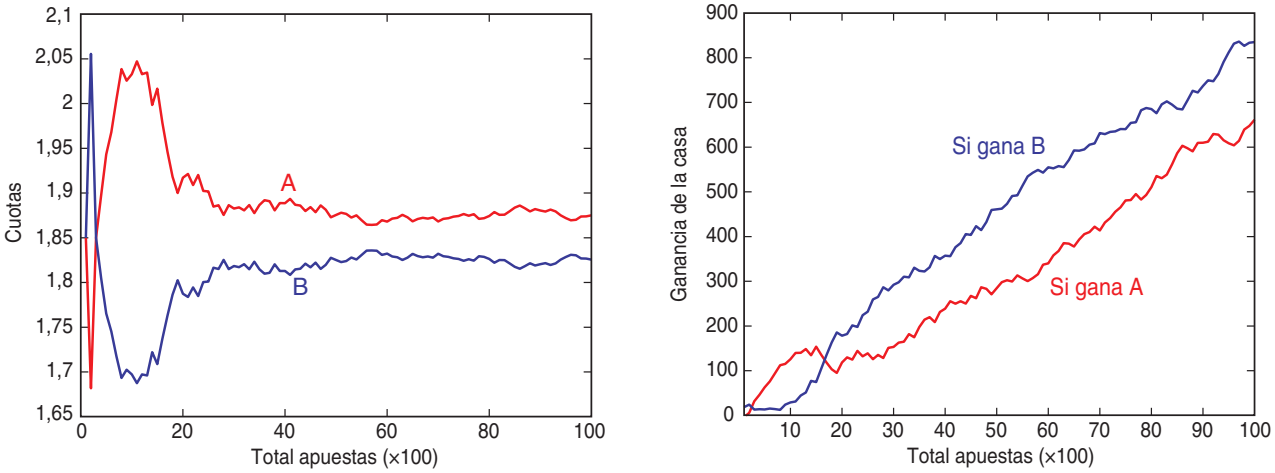
Supongamos primero que la casa no se queda ninguna comisión, sino que reparte entre los participantes todo el dinero apostado. Consideraremos únicamente apuestas con dos resultados posibles, A o B. Si se ha apostado un total de A euros a favor de A y B euros a favor de B, cuando el resultado final sea A los ganadores recibirán, por cada euro apostado, una cantidad:

$$C_A = \frac{A+B}{A} = \frac{1}{a}$$

que es precisamente la cuota de A. En esta fórmula hemos llamado a a la fracción de apuestas en favor de A, es decir, a la cantidad $A/(A+B)$. Por otro lado, la cuota de B será:

$$C_B = \frac{A+B}{B} = \frac{1}{1-a}$$

Como era de esperar, la cuota de un resultado es mayor cuanto menor es la cantidad apostada a favor de dicho resultado. Las cuotas de la tabla de nuestro ejemplo se han calculado con estas fórmulas, suponiendo que en la casa de apuestas 1 la mitad de las apuestas han sido a favor de A y la otra mitad a favor de B, es decir, $a = 1/2$, y que en la casa 2 la fracción de apuestas en favor de A es $a = 0{,}55$ y, en favor de B, 0,45. Por tanto, ¡nuestra apuesta segura es perfectamente posible!



Evolución de las cuotas (*izquierda*) y de la ganancia de la casa de apuestas (*derecha*) en función de la cantidad total apostada a lo largo del tiempo.

No se precipiten. Si la casa de apuestas se queda con una comisión o porcentaje *c* de todas las apuestas realizadas, las cuotas bajan considerablemente. Ahora serán:

$$C_A = \frac{(A+B)(1-c)}{A} = \frac{1-c}{a}\ ;\ C_B = \frac{(A+B)(1-c)}{B} = \frac{1-c}{1-a}$$

Este método de fijar las cuotas se enfrenta con un problema práctico: el apostante quiere saber lo que va a cobrar cuando apuesta y las casas garantizan el pago con las cuotas vigentes en el momento de hacer la apuesta. Para ello, actualizan las cuotas con las apuestas que van recibiendo. Inicialmente las fijan de acuerdo con la probabilidad que los expertos atribuyen a cada resultado. Si A ocurre con una probabilidad *P* y B ocurre con probabilidad 1 – *P*, las cuotas iniciales serán:

$$C_A = \frac{1-c}{P}\ ;\quad C_B = \frac{1-c}{1-P}\ ;$$

que coinciden con las anteriores si los apostantes tienen la misma percepción acerca de estas probabilidades: en ese caso apostarán a A con una probabilidad *P*, con lo que *a* será igual a *P*. Pero, obviamente, puede haber fluctuaciones. He realizado una simulación de las apuestas que recibe una casa ante un acontecimiento deportivo en el que los dos resultados tienen la misma probabilidad inicial, $P = 1/2$. El resultado puede verse en la figura 1. He supuesto que la casa actualiza las cuotas cada vez que recibe un total de 100 euros de apuestas, de los cuales una cantidad aleatoria, entre 40 y 60 euros, se inclina por A y el resto por B. La comisión de la casa es del 7,5 %, similar a la que utilizan muchas casas de Internet. En la gráfica de la izquierda se puede ver la evolución de las cuotas, mientras que en la de la derecha se representa la ganancia total de la casa según el resultado final del evento. Como vemos, las cuotas varían apreciablemente, sobre todo al principio, y la ganancia de la casa fluctúa pero es siempre positiva y tiende a lo esperado: la comisión sobre el total de apuestas, es decir, 750 euros (el 7,5 % de 10.000), aunque puede ser ligeramente inferior o superior, debido a las fluctuaciones de las apuestas, que siempre se recogen en las cuotas *a posteriori*.

Volvamos ahora a nuestro método de ganancia segura. Supongamos que las cuotas de la casa 1 son, respectivamente, $C_{1,A}$ y $C_{1,B}$, y las de la casa 2, $C_{2,A}$ y $C_{2,B}$. Si apostamos una cantidad *p* al resultado A en la casa 1 y una cantidad *q* al resultado B en la casa 2, la ganancia neta será

$pC_{1,A} - (p + q)$ si sale el resultado A
$qC_{2,B} - (p + q)$ si sale el resultado B

Para escribir la condición de ganancia segura es mejor utilizar la fracción ganada en cada apuesta que la cuota, es decir, $\alpha_1 = C_{1,A} - 1$ y $\beta_1 = C_{1,B} - 1$ y (y lo mismo para la casa de apuestas 2). En este caso, las dos ganancias serán positivas si

$$\alpha_1 p > q > \frac{p}{\beta_2}$$

Por lo tanto, basta que $\alpha_1 > 1/\beta_2$ o, equivalentemente, que $\alpha_1\ \beta_2$ sea mayor que uno, para poder hacer una apuesta completamente segura. Pero también si $\alpha_2\ \beta_1 > 1$ podemos hacer una apuesta segura, sin más que cambiar la casa 1 por la 2. Estas condiciones son fáciles de comprobar en una tabla como la de nuestro anterior ejemplo. Basta restar uno a todas las cuotas y multiplicar en cruz los valores. Si alguno de los productos es mayor que uno, la apuesta segura es posible. En la tabla del ejemplo, el producto $1 \times 1{,}25$ es mayor que uno, lo que nos ha permitido diseñar una apuesta de ganancia segura. Con un poco de álgebra y con las fórmulas anteriores, se puede demostrar que la apuesta segura es posible siempre que:

$$a_1 - a_2 > c \text{ o } a_2 - a_1 > c$$

en donde *c* es de nuevo el porcentaje de comisión (igual para las dos casas) y a_1 y a_2 son las fracciones apostadas en favor de A en la casa de apuestas 1 y 2, respectivamente. Si la comisión es nula, estas condiciones se verifican en cuanto haya la más mínima diferencia entre las apuestas realizadas en las dos casas. Cuando existe una comisión, estas condiciones no se satisfacen tan fácilmente. Para comprobarlo, podemos utilizar como ejemplo las cuotas de la figura 1. Al igual que hablamos de cuotas en distintas casas de apuestas, podemos utilizar el método con las cuotas de una misma casa en distintos momentos. Con las cuotas de la figura 1, sí es posible realizar una apuesta de ganancia segura, apostando justo en los dos momentos en donde las dos curvas alcanzan su máximo. Sin embargo, en mi simulación, aproximadamente sólo una de cada diez veces he encontrado situaciones en donde el método de ganancia segura es aplicable y siempre gracias a las grandes fluctuaciones iniciales.

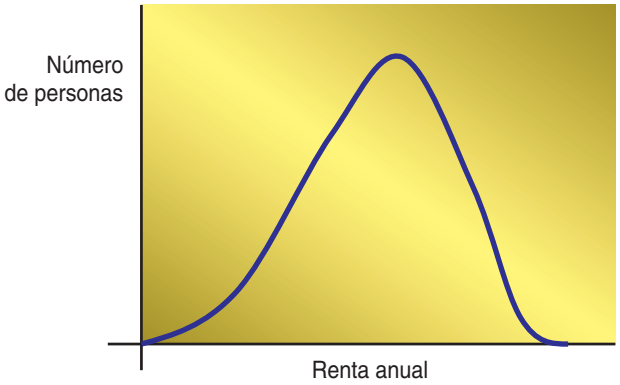
¿Es aplicable el método de ganancia segura en la práctica? Mi experiencia hasta ahora ha sido bastante desastrosa. No he sido capaz de encontrar ni una sola apuesta en la que el método sea aplicable. No sólo eso, sino que los productos cruzados suelen estar muy por debajo de 1, a pesar de que las comisiones son aparentemente bajas, en torno al 7,5 %, como en mi simulación. Sin embargo, mi búsqueda ha sido bastante limitada, porque hay muy pocas apuestas con sólo dos resultados. Por ejemplo, la mayoría de las apuestas sobre partidos del mundial de fútbol se refieren sólo a los primeros 90 minutos de juego y ofrecen los tres resultados clásicos 1-X-2. De hecho, en mi búsqueda de apuestas con ganancia segura, sólo he podido utilizar las de los resultados de los partidos de tenis de Wimbledon.

Mi impresión es que, aunque improbable, cabe encontrar alguna posibilidad de ganancia segura. Pero hay que buscar apuestas que se hayan iniciado hace poco, de modo que las cuotas fluctúen suficientemente, o también apuestas en donde haya un alto componente emocional. Por ejemplo, ante un partido de fútbol entre dos selecciones nacionales, uno debería buscar en casas de apuestas locales de cada uno de los dos países involucrados. Finalmente, convendría extender el método a tres o más resultados, puesto que la mayoría de las casas ofrecen apuestas de este tipo. No les garantizo nada, pero, mejorando el método y buscando en Internet, quizá puedan encontrar un atisbo de certidumbre en el azaroso universo de las apuestas. Si lo consiguen y se hacen millonarios, ¡acuérdense de quién les dio la idea!

Juan M.R. Parrondo

Medir la desigualdad

La economía es una disciplina muy compleja. Aunque algunos economistas mediáticos se empeñen en convencernos de que es una ciencia exacta con recetas indiscutibles (y la mayoría de las veces tendentes al neoliberalismo), lo cierto es que nadie dispone de un modelo macroeconómico perfecto, ni suficientemente contrastado. Algunos principios válidos para una época o un país pueden dejar de funcionar en otro contexto en el que sean diferentes los modos de producción, la técnica, los hábitos culturales y las motivaciones de la población. También los métodos de medición de algunos parámetros macroeconómicos, como el producto nacional bruto, están sujetos a constante revisión y mejora. Y, aun disponiendo de datos precisos y modelos fiables, ¿cuál debería ser el objetivo de la política económica? ¿Aumentar el producto nacional bruto, la cohesión social, el bienestar? Si la primera de estas variables es fácilmente cuantificable, ¿cómo podemos medir las dos últimas?

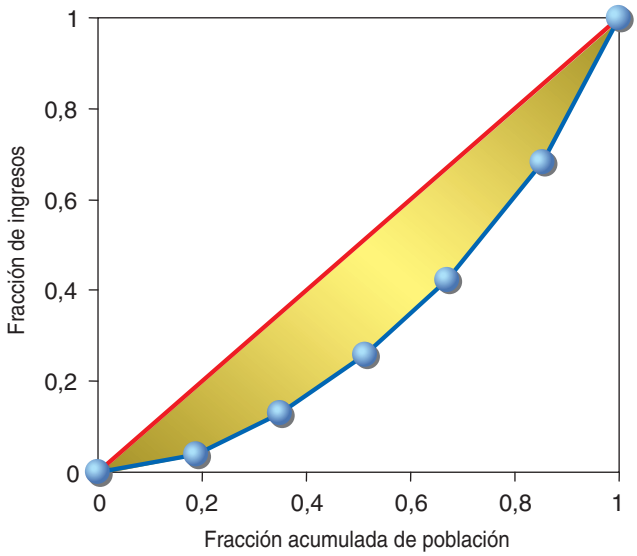


1. Posible distribución de la renta en un país.

Algunos economistas han dedicado sus esfuerzos a encontrar parámetros que cuantifiquen aspectos económicos de importancia social. Uno de ellos es la desigualdad económica, en términos de renta o en términos de patrimonio. El punto de partida en ambos casos es la distribución de renta o patrimonio, respectivamente, a lo largo de la población. En la figura 1 vemos una posible distribución de la renta anual en un país. La desigualdad económica se podría cuantificar a través de la "anchura" de la curva. Se han propuesto muchas magnitudes que miden esta anchura, como la diferencia entre el más rico y el más pobre o la dispersión de las rentas. Sin embargo, a lo largo del tiempo se ha impuesto con bastante fuerza el llamado *índice o coeficiente de Gini*, definido por el estadístico italiano Corrado Gini en un trabajo de 1912. En lugar de trabajar con la distribución de renta de la figura 1, Gini utilizó la llamada *curva de Lorenz*, en la que se representa la fracción de población que dispone de una fracción dada de la riqueza o de los ingresos totales de un país.

Para entenderlo mejor, veamos un ejemplo con los ingresos anuales de los hogares españoles en 2004, mostrados en la tabla. Se han dividido los hogares en 6 tramos de acuerdo con sus ingresos anuales. En cada tramo podemos ver el número total de hogares y el total de ingresos de dichos hogares (los datos han sido tomados del Instituto Nacional de Estadística, aunque han sido ligeramente modificados para no tener en cuenta hogares con ingresos desconocidos y para calcular los ingresos por tramo de renta). A partir de estos datos he calculado el porcentaje sobre el total, tanto de hogares como de ingresos y, finalmente, los porcentajes acumulados, es decir, la suma de porcentajes correspondientes a tramos inferiores o iguales a uno dado. Estos porcentajes acumulados, coloreados en la tabla, son los que se utilizan para dibujar la curva de Lorenz, mostrada en la figura 2 (se suelen utilizar fracciones en lugar de porcentajes, por lo que he dividido entre 100 los datos de la tabla para conseguir los puntos de la figura).

Cualquier curva de Lorenz, como le ocurre a la de la figura 2, comienza en el punto (0,0) y termina en el (1,1), y es creciente, cóncava y se encuentra siempre por debajo de la diagonal. Nos informa de qué fracción de población acumula una cierta fracción de ingresos. Por ejemplo, a partir de la tabla o de la figura 2, podemos decir que, en España, el 85,26 % de los hogares dispone del 68,36 % de los ingresos totales. En ocasiones, para indicar la desigualdad de un país, se pueden leer en la prensa afirmaciones como: "el 10 % de la población posee el 70 % de la riqueza total del país", o, equiva-



2. Curva de Lorenz de los ingresos de las familias españolas. El área amarilla, multiplicada por dos, es el índice de Gini.

3. El índice de Gini en el mundo.

lentemente, “el 90 % de la población sólo dispone del 30 % de la riqueza”. Este dato no es más que un punto en la curva de Lorenz, un punto cuya coordenada horizontal sería 0,9 y cuya coordenada vertical sería 0,3. Esta hipotética curva de Lorenz, afortunadamente muy distinta de la de nuestro país, estaría muy pegada al eje horizontal, puesto que pasa por el punto (0.9, 0.3) y luego saltaría bruscamente hasta el punto (1,1). Es una curva de alta desigualdad. Por el contrario, la diagonal roja de la figura 2 corresponde a una sociedad completamente igualitaria, en donde todos los hogares tendrían los mismos ingresos.

Pues bien, el índice de Gini, *G*, mide cuánto se separa la curva de Lorenz de esta diagonal que corresponde a la igualdad perfecta. Más concretamente, es el área comprendida entre la curva y la diagonal, es decir, el área de la región sombreada en amarillo en la figura 2, multiplicada por 2. Esta multiplicación se realiza para que el índice *G* tome valores entre 0 (igualdad absoluta) y 1 (desigualdad absoluta).

Una de las ventajas del índice de Gini frente a otros indicadores de desigualdad estriba en lo siguiente: puesto que se define en términos de fracciones de población y riqueza, es independiente de los valores totales, es decir, de la población y del producto interior bruto del país. Tiene otras propiedades matemáticas interesantes. Por ejemplo, se puede demostrar que es igual al valor medio de la diferencia relativa entre los ingresos de dos personas cualesquiera. Con otras palabras, si tomamos todos los posibles pares de personas, calculamos la diferencia de ingresos, sumamos todas las diferencias y dividimos por el número de pares, obtenemos el valor medio de la diferencia. Dividiendo por la renta per cápita, tendríamos ese valor medio en términos relativos, que coincide precisamente con el índice de Gini.

En la figura 3 podemos ver un mapa del mundo con los índices de Gini de cada país. Se puede apreciar una cierta correlación entre igualdad y desarrollo, con una excepción generalizada en los países ex comunistas, que tienden a índices de Gini más bajos, y también con notorias excepciones como el alto nivel de desigualdad de los Estados Unidos y el bajo Gini de India o algunos países africanos. En Internet pueden encontrar más datos, como la evolución del índice de Gini a lo largo del tiempo. Con ellos se puede analizar el efecto de algunas políticas sobre la desigualdad. Por ejemplo, tanto en Reino Unido como en los Estados Unidos, se aprecia un incremento de la desigualdad coincidente con las políticas económicas de Reagan y Thatcher.

El índice de Gini está sujeto a algunas ambigüedades en su cálculo, como la elección de familias o personas para dibujar la curva de Lorenz o la forma de computar sus ingresos. Quizá la base de datos más rigurosa y documentada sobre desigualdad e índices de Gini sea la del Instituto Mundial para la Investigación en Economía del Desarrollo de la Universidad de las Naciones Unidas (WIDER, www.wider.unu.edu). Otra fuente importante de datos es el Programa de Desarrollo de las Naciones Unidas (www.undp.org), en donde pueden encontrar diversos indicadores de desigualdad de todos los países

	Número de hogares (en miles)	Porcentaje de hogares	Porcentaje acumulado de hogares	Total ingresos (en millones de euros)	Porcentaje de ingresos	Porcentaje acumulado de ingresos
Hasta 9000 euros	2762	18,80	18,80	12,43	4,04	4,04
De 9000 a 14.000 euros	2376	16,18	34,98	27,32	8,88	12,91
De 14.000 a 19.000 euros	2394	16,30	51,28	39,50	12,83	25,74
De 19.000 a 25.000 euros	2317	15,77	67,05	50,97	16,56	42,30
De 25.000 a 35.000 euros	2675	18,21	85,26	80,24	26,06	68,36
Más de 35.000 euros	2164	14,74	100,00	97,40	31,64	100,00
Totales:	14.688	100,00		307,86	100,00	

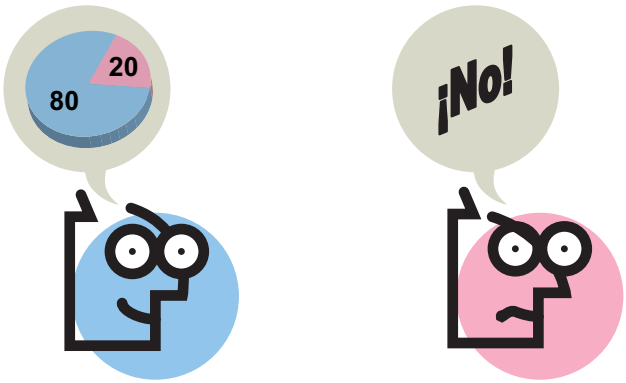
del mundo. Si bien el índice de Gini ha recibido algunas críticas y se dispone de otros indicadores alternativos, lo cierto es que hoy en día es uno de los más utilizados y se ha convertido en una herramienta econométrica imprescindible, como pueden comprobar si buscan en Internet referencias al mismo. El simple hecho de que dirija la atención de los economistas hacia el problema de la desigualdad económica lo convierte en un concepto que merece la pena difundir y estudiar con más profundidad. Aquí hemos discutido únicamente su definición, pero existen multitud de trabajos que estudian sus propiedades matemáticas, que diseñan nuevos métodos para su cálculo o tratan de incorporarlo a modelos macroeconómicos.

Juan M.R. Parrondo

El juego del ultimátum

¿Se puede cuantificar el egoísmo? En 1982, los economistas Güth, Werner, Schmittberger y Schwarze diseñaron un experimento muy sencillo (aunque no barato) que ha posibilitado el estudio cuantitativo de la cooperación y el altruismo en la conducta humana. El experimento se conoce como "juego del ultimátum" y en él participan dos jugadores, aunque cada uno desempeña un papel diferente. Uno se denomina "proponente" y el otro "contestador". El experimentador les ofrece una cantidad de dinero, pongamos 100 euros. El proponente impone el reparto que se le antoje, por ejemplo 80 euros para él y 20 para el otro jugador. Pero es este último, el contestador, quien decide aceptar o rechazar la propuesta de reparto. Si la acepta, cada uno se lleva la cantidad propuesta por el proponente. Pero si el contestador rechaza la oferta... entonces *ambos* se vuelven a casa con las manos vacías.

¿Cuál debería ser la estrategia de cada jugador si ambos actuaran racionalmente y con el único objetivo de maximizar su ganancia? En principio se podría pensar que una propuesta de reparto muy desigual, como la de la figura 1, ofenderá al contestador y será airadamente rechazada. De hecho, así ocurre en los experimentos realizados. Sin embargo, si el contestador tuviera una conducta racional y tratara únicamente de maximizar su ganancia, debería aceptar cualquier oferta, puesto que, rechazándola, estaría perdiendo dinero. A su vez, el proponente, previendo esta conducta racional, debería ofrecer el reparto más desigual posible. Si, por ejemplo, la mínima unidad monetaria de la que dispone para hacer el reparto es un euro, la estrategia racional para el proponente sería la oferta 99/1. El contestador racional, libre de cualquier condicionante emocional, debería contentarse con el euro que se le ofrece, que es en cualquier caso mejor que nada.



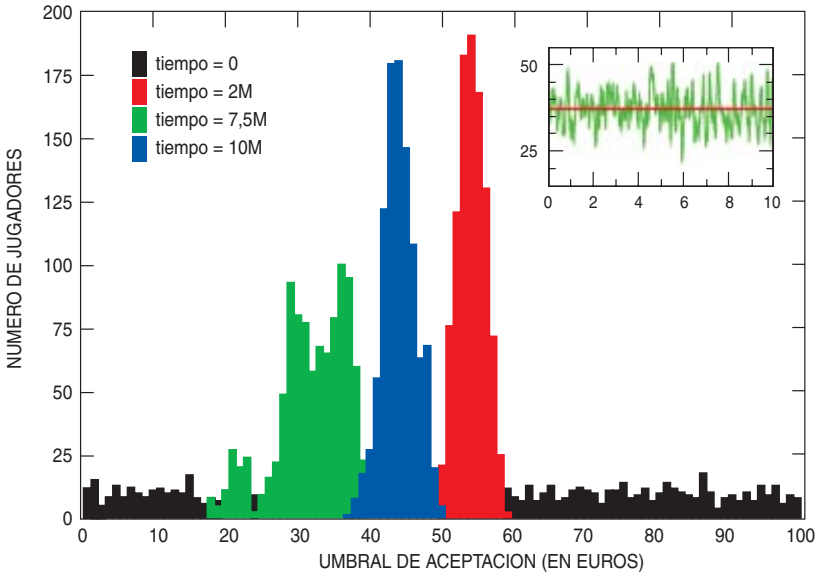
1. Un ejemplo de juego del ultimátum en el que el proponente (*izquierda*) ofrece el reparto 80/20 y el contestador (*derecha*) lo rechaza. En este caso, la ganancia de ambos jugadores es nula. El contestador ha sacrificado su ganancia de 20 euros para "castigar" el comportamiento egoísta del proponente.

Se puede demostrar que estas estrategias —proponer el reparto más desigual posible en el caso del proponente y aceptar cualquier oferta en el caso del contestador— constituyen lo que en teoría de juegos se llama *equilibrio de Nash*, uno de los conceptos básicos de la teoría de elección racional, sobre la que se basa gran parte de la economía moderna. Cualquier desviación con respecto a esta estrategia racional implica la presencia de algún tipo de componente emocional o de valoración ética.

En los últimos 20 años, se han publicado trabajos experimentales realizados con individuos de países y culturas muy diversas y con dinero real, que a veces alcanza sumas equivalentes al sueldo medio de tres meses. La mayoría de los experimentos se hacen sin que los jugadores se vean las caras o sepan quién es su contrincante y de modo que cada individuo participe una sola vez. Se elimina así la posibilidad de que los jugadores estén influidos por la identidad del oponente o que elaboren algún tipo de estrategia de intercambio. Por supuesto, todos los participantes conocen las reglas del juego y el dinero a repartir.

Los resultados se alejan mucho del comportamiento "racional". Los contestadores no aceptan cualquier cosa: suelen rechazar ofertas muy desiguales. Es decir, son capaces de sacrificar ganancias significativas con tal de castigar a un proponente excesivamente egoísta. Por su parte, el comportamiento típico de los proponentes es ofrecer repartos equitativos, 50/50, o ligeramente favorables. En su variante *juego del dictador*, el contestador es absolutamente pasivo y no tiene opción de rechazar la oferta; éstas se hacen más desiguales, pero sin alejarse mucho de la opción equitativa. Esto indica que la motivación de los proponentes para ofrecer repartos equitativos en el juego del ultimátum es doble: por un lado, un cierto sentido ético o de empatía con su contrincante y, por otro, el temor a que el contestador rechace una oferta no equitativa. Así se confirma en entrevistas realizadas al acabar el juego.

Estos resultados son bastante generales, aunque hay ciertas variaciones culturales. Por ejemplo, los Machiguenga del Amazonas peruano suelen ofrecer al contestador cantidades muy bajas, en torno al 26 % del total, y éste suele aceptarlas. Para ellos, el reparto de papeles en el juego es parte del mismo: el contestador no considera una oferta desigual como una muestra de egoísmo por parte del proponente, sino como resultado de su propia mala suerte. Por otro lado, entre los pastores Sukuma de Tanzania la media de las ofertas alcanza un 61 % del total a repartir. Brian Paciotti, de la Universidad de California en Davis, ha relacionado este comportamiento generoso con ciertas aptitudes de los Sukuma para la



2. Evolución del umbral de aceptación en el modelo de Sánchez y Cuesta para mutaciones rápidas (s = 1), inicialmente (*en negro*) y tras 2 (*rojo*), 7,5 (*verde*) y 10 (*azul*) millones de turnos. La simulación se realizó con 1000 participantes que juegan al ultimátum con 100 euros por turno. En el recuadro interior (*en verde*) se puede ver la evolución del umbral medio en función del tiempo (en millones de turnos). [Tomada del *Journal of Theoretical Biology*, vol. 235, 2005.]

organización social. Como en todos los estudios antropológicos, los investigadores tienen que ser muy cuidadosos para no influir inconscientemente sobre los jugadores. Incluso la forma de explicar el juego puede dar lugar a comportamientos distintos entre un experimento y otro. Aun así, el aspecto cuantitativo del juego del ultimátum lo convierte en una herramienta útil para caracterizar la propensión a la cooperación o el sentido de justicia en cada cultura.

Recientemente se ha acuñado el término *neuroeconomía* para referirse a un nuevo campo de investigación que trata de encontrar y estudiar mecanismos de toma de decisiones con componentes no racionales. En la neuroeconomía convergen investigaciones muy variadas: desde los intentos de economistas de incluir motivaciones emocionales en la función de utilidad hasta estudios neurológicos. Por ejemplo, Alan G. Sanfey y sus colaboradores de la Universidad de Princeton han analizado, mediante técnicas de resonancia magnética, qué zonas del cerebro se activan en los participantes en el juego del ultimátum. El resultado fue el esperado: el rechazo viene siempre acompañado de la activación de la ínsula anterior, una zona asociada con distintas emociones negativas, como el asco, mientras que las ofertas muy desiguales activan en el proponente el córtex cingulado anterior, que suele estar relacionado con conflictos cognitivos.

El juego también se ha utilizado recientemente para desentrañar el origen evolutivo de la cooperación. Angel Sánchez y José Cuesta, de la Universidad Carlos III de Madrid, han diseñado un modelo sencillo de evolución de individuos que se enfrentan unos a otros en el juego del ultimátum repartiendo M euros en cada turno. Cada jugador está caracterizado por el capital que ha acumulado Xi y por un umbral u_i entre 0 y M, que es la cantidad mínima que está dispuesto a aceptar. En una primera versión del modelo, cada jugador "supone" que su contrincante piensa como él y, por tanto, le ofrece el reparto: $M - u_i$ (para el proponente), u_i (para el contestador). En cada turno se eligen al azar dos jugadores, i y j, y se les asigna, también al azar, el papel de proponente y contestador. Si, por ejemplo, i resulta ser el proponente y $u_i \geq u_J$, entonces el resultado del juego es:

$$X_j \rightarrow X_j + M - u_i$$
$$X_j \rightarrow X_j + u_i$$

Por otro lado, si $u_i \leq u_J$, j rechaza la oferta de i y los capitales de los dos jugadores no varían. El modelo incorpora el efecto de la selección natural eliminando, cada cierto número s de turnos, al jugador con menos capital y sustituyéndolo por un "hijo" del jugador con más capital. El umbral de este hijo será similar al de su padre. Más concretamente, si el jugador padre tiene un umbral u_i, el del hijo puede ser u_i, $u_i + 1$ o $u_i - 1$, ocurriendo cada una de estas posibilidades con probabilidad 1/3. Con ello se introducen mutaciones en el sistema, así como la "selección natural" de los jugadores con más éxito.

El resultado de la simulación del modelo con 1000 jugadores y $M = 100$ puede verse en la figura 2, para el caso en el que la selección (reemplazo del peor jugador por un descendiente del mejor jugador) se realiza en cada turno (s = 1). Después de un gran número de turnos (suficiente como para que todos jueguen con todos varias veces), los umbrales se agrupan en torno a un cierto valor que cambia con el tiempo, variando entre los 25 y 50 euros. Para mutaciones más lentas, s = 10.000, se alcanza una distribución estacionaria muy concentrada en torno a los 47 euros. Es curioso que los resultados de la simulación sean tan cercanos a los reales. Sánchez y Cuesta no pretenden dar una explicación evolutiva del comportamiento humano ante el juego y aún menos una explicación cuantitativa, puesto que el juego del ultimátum es una construcción artificial a la que no hemos estado sometidos los humanos a lo largo de la evolución de nuestra especie (aunque quizá haya situaciones "naturales" formalmente idénticas). Sin embargo, la simulación muestra que la selección natural puede dar lugar a comportamientos cooperativos, sin necesidad de memoria y simplemente a través de la interacción entre parejas de individuos.

Si quieren saber más sobre las matemáticas de la cooperación, pueden consultar el excelente artículo de revisión de Angel Sánchez en el número del pasado mes de junio de *Matematicalia* (www.matematicalia.net), la revista digital de divulgación matemática de la Real Sociedad Matemática Española.

parr@seneca.fis.ucm.es

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Los dados misteriosos y la razón áurea

— Los dados no engañan nunca, Simplicio. Una vez que salen de la mano, no obedecen voluntad alguna, salvo la de la fortuna. ¿No quieres probar si hoy te sonríe?

— Ya me has engañado muchas veces, Malaspina. Y aún ahora me vienes con unos dados endiablados, cada uno distinto del otro y con números que no había visto jamás. El primero tiene en sus seis caras los números 1-1-5-5-5-5, el segundo 3-3-3-4-4-4 y el tercero 2-2-2-6-6-6. ¿Cómo quieres que me atreva a jugar contigo si seguramente habrás compuesto estos dados para dejarme sin blanca?

— Yo te aseguro que ninguno de ellos es mejor que el otro, descreído. Y para demostrártelo, te dejaré elegir primero. Tú jugarás con el dado que desees y yo con alguno de los dos que rechaces. Cada uno tirará su dado y el que tenga la jugada menor pagará un escudo al otro.

— Esta parece una ventaja considerable. Déjame ver... elijo el tercero de los dados, que al menos tiene un par de seises.

— Sabia elección, Simplicio. Yo me quedaré con el segundo, por quedarme con alguno...

Después de cien tiradas, Simplicio había perdido más de 20 escudos y comenzaba a arrepentirse de su elección:

— De poco valen mis dos seises, si todos tus números ganan a mis cuatro doses. Me parece que me has vuelto a engañar. ¡No quiero seguir jugando!

— Toma entonces mi dado, Simplicio y yo tomaré el que nadie quiso en un principio: el primero.

— Con este trueque sí probaré fortuna.

Pero en cien nuevas tiradas, Simplicio volvía a perder otros 20 escudos:

— ¡Sí que es mala mi suerte! Dame ahora ese dado con el que juegas, que parece ser el mejor de todos.

— Tómalo, Simplicio. Yo me quedaré con el tercero, el que tan mala suerte te dio en las primeras cien jugadas.

Y, de nuevo, Malaspina logró vaciar los bolsillos de Simplicio, ganándole esta vez 10 escudos en otras cien jugadas.

— No importa el dado con el que juegue —se lamentaba el perdedor—, la suerte no está hoy de mi lado. Me da en la nariz que alguna trampa has escondido entre todos estos números, pero no acierto a descubrirla.

En efecto, los dados de Malaspina tienen una propiedad desconcertante, capaz de confundir a un jugador ingenuo. Ninguno de ellos es el mejor, el primero gana al segundo con bastante probabilidad, el segundo al tercero y el tercero al primero. Calcular estas probabilidades no es complicado. En la figura 1 muestro tres tablas, correspondientes a los tres posibles enfrentamientos entre los dados. He sombreado en color azul las casillas en las que el dado del eje vertical gana al del eje horizontal.

Contando las casillas azules y dividiendo por el número total de casillas, que es 36, obtenemos cada probabilidad. La de que el primer dado (A) gane al segundo (B), será $24/36 = 2/3 = 0{,}6666...$, la de que B gane a C también 24/36, y la de que C gane a A es $20/36 = 5/9 = 0{,}5555...$ Las tres son mayores que 1/2 y es por ello por lo que Malaspina siempre gana, al cabo de un gran número de jugadas, al pobre Simplicio. Estamos ante una nueva paradoja de intransitividad, similar a algunas de las que ya hemos discutido en esta sección (*Paradojas democráticas*, marzo de 2002, y *Quien ríe el último...*, noviembre de 2005).

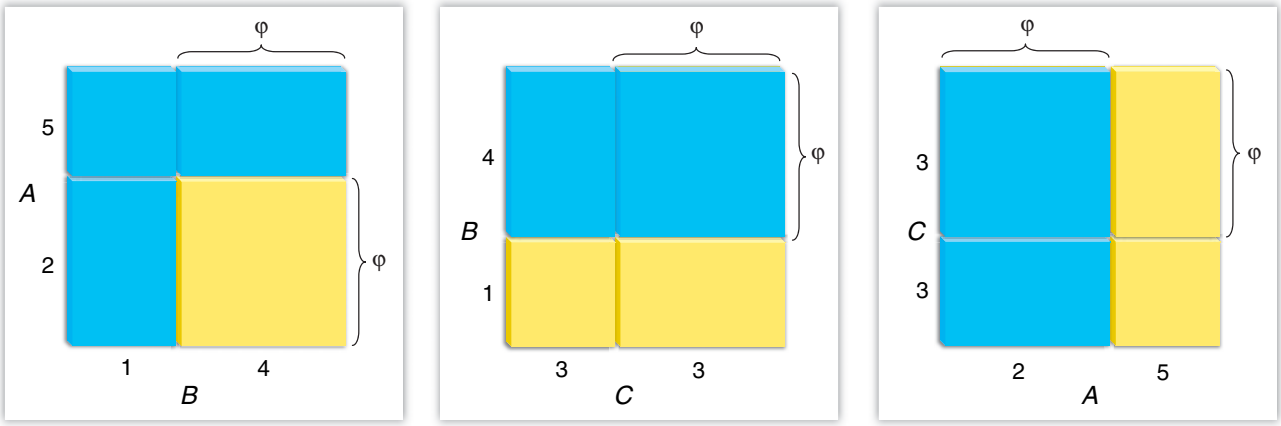
En esta ocasión, nuestros dados son un caso particular de la llamada *paradoja de Steinhaus-Trybula*, dos matemáticos que analizaron por primera vez la posibilidad de que, dadas tres cantidades aleatorias *A*, *B* y *C*, aconteciera

1. Diagramas para el cálculo de probabilidades de ganar con los dados misteriosos, A, B y C. Las casillas azules muestran las situaciones en las que el dado del eje vertical gana al del eje horizontal.

A \ B	3	3	3	4	4	4
5						
5						
5						
5						
1						
1						

B \ C	2	2	2	2	6	6
4						
4						
4						
3						
3						
3						

C \ A	1	1	5	5	5	5
6						
6						
2						
2						
2						
2						



2. Probabilidades en los enfrentamientos de tres "dados" A, B, y C. En cada enfrentamiento, el área de la región azul indica la probabilidad de que el dado vertical gane al horizontal. Las tres probabilidades de ganar son iguales a la razón áurea φ. El lado del cuadrado completo es 1 en los tres casos.

que las tres probabilidades $P(A>B)$, $P(B>C)$ y $P(C>A)$ fueran mayores que 1/2. La paradoja puede dar lugar a situaciones sorprendentes no sólo en el juego. Por ejemplo, si las cantidades *A*, *B* y *C* fuesen los tiempos de llegada de tres autobuses a una parada, sería más probable que llegase A antes que B, también sería más probable que B llegase antes que C y, finalmente, ¡lo más probable sería que C llegase antes que B! Aunque pueda parecer absurdo, el cálculo de probabilidades muestra que esta situación es perfectamente posible.

Trybula demostró más tarde un teorema bastante curioso en el que, inesperadamente, aparece una vieja amiga de la que también hemos hablado aquí: la razón áurea, $\varphi = \sqrt{5} - 1/2 = 0{,}618...$ (*El número de oro*, enero de 2003). El teorema afirma que, siempre que se dé la paradoja de intransitividad, al menos una de las tres probabilidades $P(A>B)$, $P(B>C)$ o $P(C>A)$ tiene que ser menor o igual que φ.

No podemos hacer aquí una demostración rigurosa del teorema, pero sí esbozarla para ver cómo puede aparecer en este contexto la razón áurea. Para ello, supongamos que cada una de las cantidades aleatorias, *A*, *B* y *C*, sólo puede tomar dos valores de entre los números 1, 2, 3, 4 y 5. Una forma de asignar probabilidades a esos valores viene dada por la siguiente tabla:

Valor	1	2	3	4	5
Probabilidad	1 – pB	1 – pA	1	pB	pA
Cantidad aleatoria	B	A	C	B	A

Por ejemplo, la cantidad *A* sólo puede tomar los valores 2 y 5 y lo hace con probabilidades, $1 - p_A$ y p_A, respectivamente. Este reparto de los posibles valores entre las tres cantidades aleatorias es bastante igualitario, puesto que se van repartiendo los cinco valores de mayor a menor, como en la elección del equipo de fútbol en el patio de un colegio. Uno de los pasos cruciales y más ingeniosos de la prueba del teorema de Trybula es precisamente demostrar que este reparto es el que maximiza las probabilidades de que una cantidad sea mayor que otra, manteniendo la paradoja. A partir de esta tabla pueden encontrarse las tres probabilidades de ganar en cada enfrentamiento:

$$P(A>B) = p_A + (1-p_A)(1-p_B)$$
$$P(B>C) = p_B$$
$$P(C>A) = 1 - p_A$$

Otro paso sutil de la demostración consiste en probar que, en el reparto que maximiza las probabilidades (más precisamente: el reparto que maximiza la más pequeña de las tres probabilidades), éstas tienen que ser iguales. La razón es que, si no lo fueran, siempre podríamos aumentar alguna de ellas ligeramente, variando alguno de los dos parámetros p_A o p_B. Una vez que se llega a esta conclusión, lo que queda es un poco de álgebra: se igualan las tres probabilidades y, de las dos ecuaciones resultantes, se obtienen p_A y p_B, que resultan ser: $p_A = 1 - f$ y $p_B = f$. Finalmente, se calculan las probabilidades con esta solución y obtenemos que las tres valen precisamente φ.

Trybula diseñó además un ejemplo en el que este áureo valor máximo se puede alcanzar. De hecho, la propia demostración que hemos esbozado aquí conduce inmediatamente al ejemplo. En la figura 2 se ve gráficamente el cálculo de probabilidades. Como en las tablas de la figura 1, las áreas azules indican la probabilidad de que la cantidad aleatoria del eje vertical gane a la del eje horizontal. Las tres áreas valen exactamente φ, gracias a una de las propiedades matemáticas del número de oro: $\varphi = 1 - \varphi^2$.

Les dejo con otros dados paradójicos, tan curiosos o más que los anteriores. En el A las caras contienen los números 1-5-5-5-5-7, en el B, 2-4-4-6-6-6, y en el C, 3-3-3-3-8-8. En este caso, A gana a C y B también gana a C. Pues bien, a pesar de que, en apariencia, C es el peor de los tres dados, si jugamos con todos ellos a la vez, es decir, tirando los tres y otorgando la victoria al de mayor puntuación, ¡el que tiene la probabilidad más alta de ganar es C! Si trasladamos la paradoja a la llegada de los autobuses, significaría que lo más probable es que A llegue antes que C, que B llegue antes que C, pero el que tiene más probabilidades de llegar primero de los tres es C. El cálculo de las probabilidades es bastante sencillo, aunque el resultado parezca increíble. Todas estas paradojas tienen además un interés no sólo recreativo o académico, ya que ponen en jaque la suposición básica de la teoría de la utilidad: que somos capaces de ordenar nuestras preferencias de modo lineal.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La joya oculta

Algunos piensan que la matemática es una construcción del ser humano. Otros defienden lo contrario: que las verdades matemáticas están —desde siempre— "ahí afuera", esperando ser descubiertas. La historia de la matemática da y quita la razón a ambas posturas: en ocasiones los matemáticos construyen realidades matemáticas como si fueran artefactos, pero muchas otras veces —quizá las más— parecen desorientados exploradores de un territorio virgen. A este segundo caso pertenece la invención de los logaritmos por John Napier (o Neper) en 1614. Napier se enfrentó a un problema que podía resolverse de dos maneras. Eligió la más complicada (aunque luego rectificaría con la ayuda de Henry Briggs) y encontró, sin darse cuenta, una auténtica joya, el número *e*. Napier lo tuvo en sus manos, sin percatarse de su inmenso valor. De hecho, el número *e* estuvo oculto en el corazón de los logaritmos *neperianos* originales (que son distintos de los neperianos modernos) hasta que Euler, un siglo después, lo hiciera brillar con todo su esplendor.

Los logaritmos nacieron como una herramienta, un truco, para calcular rápidamente proporciones, productos, raíces y potencias. Fueron descubiertos a finales del siglo XVI independientemente por Napier y por Burgi, quien, sin embargo, no dio la publicidad debida a su hallazgo. Napier sí publicó sus tablas en latín en 1614: *Mirifici logarithmorum canonis descriptio*, traducida al inglés en 1616 con el título: *A description of the admirable table of logarithmes*. Después de morir Napier, se publicó en 1619 su *Mirifici logarithmorum canonis constructio*, en donde explica en detalle cómo había construido las tablas de la *Descriptio* de 1614. Burgi y Napier se basaron en una idea que se remonta al menos a Arquímedes. Consideremos la siguiente tabla de potencias de 2:

n	1	2	3	4	5	6	7	8	9	10	11	12
2^n	2	4	8	16	32	64	128	256	512	1024	2048	4096

Multiplicar y dividir los números de la segunda fila de esta tabla es muy sencillo. Por ejemplo, para calcular 16 × 64 basta sumar los números correspondientes de la primera fila: 16 es la cuarta potencia de 2 y 64 la sexta; sumamos entonces 4 y 6, y obtenemos 10; el resultado del producto es entonces la potencia décima, es decir, 16 × 64 = 1024. Hoy llamamos "logaritmos en base 2" a los números de la primera fila. Elevar al cuadrado un número es también muy sencillo: basta multiplicar por dos su logaritmo. Finalmente, las raíces cuadradas o cúbicas se calculan dividiendo el logaritmo por dos o por tres, respectivamente. Cálculos complicados como:

$$\frac{\sqrt[2]{1024}\times 4^5}{\sqrt[3]{64}\times 256}$$

se simplifican enormemente con la tabla de logaritmos: el logaritmo de 1024 es 10, el de 4 es 2, el de 64 es 6, el de 256 es 8. La operación anterior, con logaritmos, se reduce a 10/2 + 2 × 5 − 6/3 − 8 = 5, que es el logaritmo de 32. Por tanto, el resultado de la expresión anterior es 32.

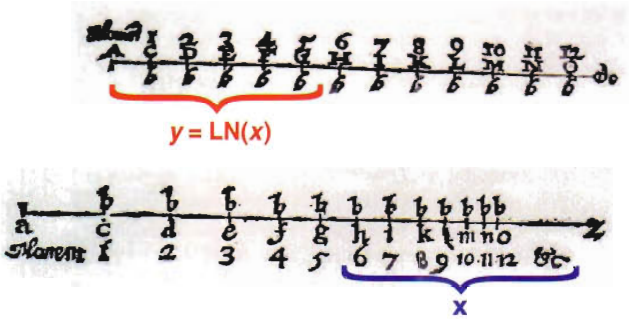
La idea es muy sugerente, pero tiene limitaciones obvias. La primera es que sólo podemos aplicarla a cálculos con los números que aparecen en la tabla, es decir, con potencias de 2. En segundo lugar, ¿qué ocurre si después de operar con los logaritmos el resultado es negativo o fraccionario? La solución más sencilla es ampliar la noción de potencia a exponentes negativos y fraccionarios. La regla de cálculo de logaritmos nos dice que dividir el logaritmo entre 2 es equivalente a extraer su raíz cuadrada. Análogamente, podemos definir, por ejemplo, $2^{1/2} = \sqrt{2} = 1{,}414...$ También de acuerdo con las reglas anteriores, los exponentes negativos tienen que definirse en la forma: $2^{-n} = 1/2^n$.

Esta es la solución sencilla —al menos a nuestros ojos— de las limitaciones de la idea original de Arquímedes. Otra posibilidad, bastante más complicada, es escoger una base próxima a la unidad, que daría lugar a potencias muy cercanas entre sí. Probablemente Napier partió de esta idea, pero la refinó utilizando un símil cinemático. Sabía que la esencia del truco de Arquímedes era comparar una progresión geométrica con una aritmética, pero necesitaba progresiones que recorrieran todos los números posibles.

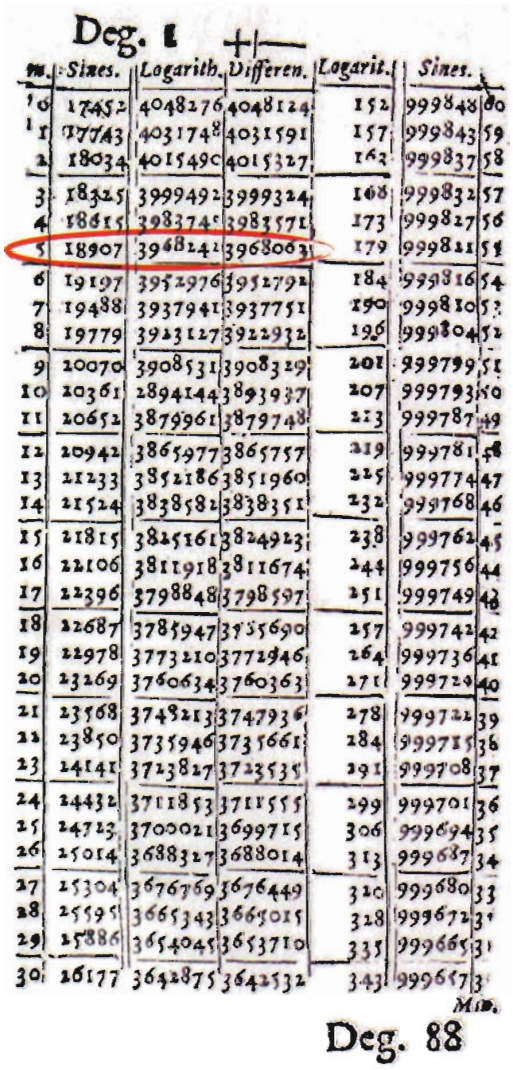
Imaginó entonces dos puntos moviéndose sobre sendas líneas. El primero con velocidad constante v_0 y el segundo con una velocidad proporcional a la distancia x que le separa del fin de la línea. En la figura 1 vemos los grabados originales de Napier. En la línea superior, el móvil se mueve a velocidad constante y los puntos CDEFG, etc. muestran su posición a intervalos regulares de tiempo. En la línea inferior, otro móvil se mueve con una velocidad decreciente, proporcional x: recorre los puntos *acdefg*, etc., a intervalos regulares de tiempo; pero, como la velocidad disminuye paulatinamente, estos puntos están ahora cada vez más cercanos entre sí. Los puntos de la línea superior desempeñan la función de la progresión aritmética y los de la inferior el de la progresión geométrica: los intervalos AB, AC, AD, etc. difieren en una cantidad constante, mientras que en la línea inferior son los cocientes entre los intervalos *ab*, *ac*, *ad*, etc. los que son constantes. Finalmente, Napier define el logaritmo $y = \mathrm{LN}(x)$ de un número x de la siguiente forma: cuando el punto de la línea inferior se encuentra a una distancia x del final de la misma, el punto de la línea superior ha recorrido una distancia $y = \mathrm{LN}(x)$.

Los logaritmos neperianos modernos son diferentes. Hoy sabemos que el punto de la línea superior se mueve siguiendo la ecuación $y(t) = v_0 t$, mientras que el inferior, cuya velocidad en un instante t es $v(t) = v_0 x(t)/w$, obedece a la ecuación $x(t) = we^{-v_0 t/w}$. De ello se deduce la relación entre los neperianos modernos y los antiguos:

$$LN(x) = -w \ln \frac{x}{w}$$



1. Los dos movimientos de Napier: en el de arriba el móvil avanza con velocidad constante y en el de abajo con velocidad proporcional a la distancia x que separa el punto del final z de la línea. Para una distancia x, Napier define el logaritmo de x, LN(x) como la distancia y que ha recorrido el móvil de la línea superior.



2. Extracto de una de las tablas de logaritmos originales de Napier.

Napier tomó como longitud total de la línea *az* la cantidad $w = 10^6$. En la figura 2 tenemos una de las tablas de la *Descriptio,* cuyos valores son sorprendentemente precisos. Por ejemplo, para $x = 18.907$, una calculadora moderna nos da LN(18.907) = 3.968.223,055, que difiere del calculado por Napier, 3.968.242, en sólo 19 unidades en casi cuatro millones.

Con la definición de Napier, si dos parejas de puntos del movimiento de la línea inferior guardan la misma proporción:

$$\frac{x_1}{x_2} = \frac{x_3}{x_4},$$

entonces los puntos correspondientes en la línea superior, es decir, sus logaritmos, tendrán incrementos iguales:

$$LN(x_1) - LN(x_2) = LN(x_3) - LN(x_4)$$

y esta propiedad es la clave para la función última de los logaritmos: convertir productos y cocientes en sumas y restas.

¿Cómo pudo calcular Napier sus logaritmos con esta increíble precisión? Utilizó algo que hoy llamamos "integración numérica de una ecuación diferencial", pero con trucos bastante ingeniosos que le permitieron culminar la tarea, aunque empleara en ello unos veinte años. Comenzó calculando la posición x del móvil de la línea inferior con pequeños intervalos de tiempo. Si la velocidad inicial es $v_0 = 1$, una estimación de la posición del móvil al cabo de un segundo será $x(1) = w - 1 = w(1 - 1/w)$. Al cabo de 2 segundos, $x(2) = (w - 1) - (w - 1)/w = w(1 - 1/w)^2$ (recordemos que la velocidad en cada instante es x/w). Es decir, podemos encontrar la posición en el tiempo $x(t + 1)$ sin más que sustraer a $x(t)$ su 10^6-ava parte: $x(t + 1) = x(t)(1 - 1/w)$. En los instantes iniciales del movimiento, este cálculo hace descender x prácticamente en intervalos de una unidad. Sin embargo, para x menores, el descenso es mucho más lento. Por ejemplo, si $x(t) = 1000$, $x(t + 1) = 999{,}999$. Napier habría necesitado millones de operaciones para construir sus tablas con intervalos de un segundo. Halló entonces el modo de simplificar el cálculo utilizando intervalos de tiempo más largos cuando la velocidad es tan baja, que el móvil apenas varía su posición en un segundo. Este es el método de paso de tiempo variable, aplicado hoy en algunas simulaciones numéricas de sistemas dinámicos. Napier calculó tablas con diferentes pasos de tiempo y luego las refinó y combinó por medio de interpolaciones y otros trucos bastante ingeniosos.

La mayoría de sus argumentos se basaban en la intuición que le proporcionó el símil cinemático. Napier encontró sus logaritmos, íntimamente relacionados con los neperianos modernos, sin conocer la base de estos últimos, el número e. Sin embargo, el número e estaba en el corazón de su modelo cinemático y también en sus cálculos, puesto que utilizó constantemente expresiones del tipo $(1 + 1/w)^k$ con w igual a un número muy grande (10^6 y, originalmente, 10^7). La definición moderna del número e emplea este tipo de potencias: es un cierto tipo de límite de números infinitamente cercanos a 1 elevados a exponentes infinitamente grandes.

parr@seneca.fis.ucm.es

Juan M.R. Parrondo

Los logaritmos de Briggs

Verano de 1615. El matemático Henry Briggs cruza Gran Bretaña en diligencia. Hace meses cayó en sus manos el recién publicado *Mirifici logarithmorum canonis descriptio* de John Napier. Supo entonces que se hallaba ante una obra de importancia histórica. Ha escrito varias cartas al autor del tratado, alabando su trabajo y proponiendo algunas mejoras. Napier está de acuerdo con ellas, pero a sus 65 años no se ve con fuerzas para llevarlas a cabo. Briggs, catedrático del Gresham College de Londres, donde años después se originaría la Royal Society, se dirige hacia Edimburgo para conocer al maestro.

Llega con bastante retraso a la cita. Llama a la puerta de la casa de Napier y, según William Lilly, los dos hombres pasan más de un cuarto de hora contemplándose con admiración y sin pronunciar palabra. Finalmente, Briggs dice: "Señor, he realizado este largo viaje con el propósito de veros en persona y conocer por qué suerte de ingenio o inteligencia vinisteis a ser el primero en concebir ésta la más excelente ayuda a la astronomía: los logaritmos."

El mes pasado reconstruimos con algún detalle el nacimiento de los logaritmos y el trabajo de Napier. La idea fundamental, que se remonta a Arquímedes, consiste en relacionar una lista de números en progresión geométrica (los elementos consecutivos tienen cocientes iguales) con otra en progresión aritmética (los elementos consecutivos tienen diferencias iguales). Lo que Briggs le propuso a Napier en aquella visita fue utilizar como base para los logaritmos el número 10. Esta idea les condujo a la moderna definición de logaritmo en base 10: el logaritmo de un número x es el exponente y al que hay que elevar 10 para obtener x, es decir, y es el logaritmo de x si $x = 10^y$.

Después del encuentro de 1615, Briggs trabajó sobre estas ideas hasta publicar en 1624 la *Arithmetica Logarithmica* (en latín), la primera tabla de logaritmos en base 10 de la historia, calculados con tal precisión que apenas tuvo que ser corregida en los cuatro siglos siguientes. Como en el caso de Napier, lo más interesante del trabajo de Briggs fue el ingenio desplegado para reducir en lo posible los cálculos necesarios para la elaboración de la tabla. También Briggs descubrió trucos que vislumbraban la que luego sería la herramienta más potente de la matemática: el cálculo diferencial, desarrollado por Newton y Leibniz a finales del siglo XVII.

El enfoque de Briggs, sin embargo, fue distinto del seguido por Napier. Si éste utilizó un símil geométrico para definir sus logaritmos, Briggs se basó únicamente en operaciones algebraicas. Su método radica en un ingenioso truco para calcular logaritmos de números muy cercanos a la unidad.

Briggs comienza calculando la raíz cuadrada de 10, luego la raíz cuadrada del resultado, y así sucesivamente hasta 54 veces. Hoy se pueden hacer estas operaciones con una calculadora de bolsillo en unos pocos segundos,

1. Las 54 raíces cuadradas sucesivas de 10, tal y como aparecen en la *Arithmetica*.

pulsando la tecla raíz 54 veces. Si su calculadora no tiene suficiente precisión, verán que el resultado final es 1. Sin embargo, el valor real no es exactamente la unidad. Briggs realizó el cálculo con 30 cifras decimales, tal y como vemos en la figura 1, tomada de la *Arithmetica*. El último número es un 1 seguido de 16 ceros y de 14 cifras más. Es en realidad la raíz 2^{54}-ava de 10, o $10^{1/2^{54}}$, puesto que extraer una raíz cuadrada es equivalente a dividir entre dos el exponente de una potencia: $\sqrt{10^n} = 10^{n/2}$. En la columna de la izquierda de la tabla se muestran las raíces; en la de la derecha, Briggs escribe los exponentes correspondientes, es decir, los logaritmos de los números de la izquierda. Si en la columna de la izquierda cada número es la raíz cuadrada del anterior, en la columna de la derecha cada número o logaritmo es la mitad del anterior, puesto que así cambian los exponentes cuando se extrae la raíz cuadrada de una potencia.

Los últimos números de la tabla tienen una curiosa propiedad: si nos olvidamos del 1 y sólo atendemos a la parte decimal, vemos que el último número es exactamente la mitad del penúltimo, a pesar de que aquél es la raíz cuadrada de éste. En notación moderna:

$$\sqrt{1+\varepsilon} \approx 1 + \frac{\varepsilon}{2}$$

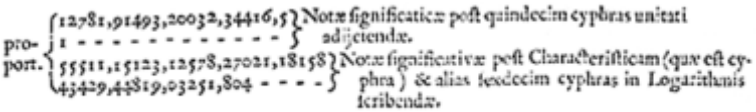
Como vemos en esta fórmula, la parte decimal ε de un número cercano a la unidad se divide entre dos cuando extraemos la raíz cuadrada. Se trata en realidad de una aproximación. Hoy sabemos que, el error en la fórmula anterior es del orden de ε^2, que en el caso de los últimos números de la tabla de Briggs, afecta precisamente a la 30-ava cifra decimal.

Una vez convencido de que, en las cercanías de la unidad, números y logaritmos se comportan igual, se pueden aplicar simples reglas de tres para obtener el logaritmo de cualquier número cercano a 1. Para hacerlo de forma sistemática, Briggs calculó el logaritmo de $1 + 10^{-16}$, es decir, de un 1 seguido de la coma decimal, de 15 ceros y de un 1. La regla de tres, tal y como la describe Briggs, se puede ver en la figura 2. En realidad, lo que está haciendo Briggs con estas reglas de tres es utilizar la *aproximación lineal* del logaritmo, que hoy escribimos:

$$\log_{10}(1 + \varepsilon) \approx k\varepsilon$$

Si aplicamos esta fórmula a la raíz 2^{54}-ava de 10, podemos obtener k dividiendo el último número de la segunda columna por los decimales del último de la primera columna de la figura 1. Esto es lo que se hace en el pasaje de la *Arithmetica* presentado en la figura 2. El resultado es k = 0,43429448190325. Hoy sabemos que k es en realidad la derivada del logaritmo en base 10, es decir, $k = \log_{10}(e) = 1/\ln(10)$, cuyo valor es el calculado por Briggs con una exactitud de 14 cifras decimales. El número e, como en el caso de Napier, estaba presente en los cálculos de Briggs sin que tampoco fuera consciente de la importancia que tendría después en la matemática y la física.

Con la aproximación lineal, a la que Briggs llamó la *regla de oro*, podía calcular mediante reglas de tres el



2. La regla de oro de Brigss aplicada al cálculo de $\log_{10}(e)$.

logaritmo de cualquier número suficientemente próximo a 1. El problema es que necesitaba que fueran extremadamente próximos a 1, más en concreto, que fueran un 1 seguido de 15 ceros decimales. Esta es una limitación considerable. Sin embargo, la idea de Briggs fue "llevar" cualquier número hacia las proximidades de la unidad calculando raíces cuadradas, igual que había hecho con el 10. Es una tarea considerable, pero encontró muchos trucos para hacerla más viable. Por ejemplo, para calcular el logaritmo de 2, primero lo elevó a la décima potencia, cuyo resultado es 1024, y luego lo dividió por 1000, obteniendo 1,024. Con ello, le bastó calcular 47 raíces cuadradas para obtener 1 y 15 ceros después de la coma decimal. Después de esos 15 ceros obtuvo 16851605705394977. Aplicando la regla de oro, calculó el logaritmo *l* de este número y luego deshizo todas las operaciones en el campo de los logaritmos: multiplicó *l* por 2^{47} para "deshacer" las 47 raíces cuadradas, al resultado le sumó 3, que es el logaritmo de 1000 y el resultado lo dividió entre 10 para "deshacer" la potencia inicial. El resultado final fue:

$$\log_{10}(2) = 0{,}301029995663981195$$

y, como pueden comprobar con una calculadora suficientemente potente, las 18 cifras que Briggs obtuvo son exactas. Utilizando estos trucos y otros métodos de interpolación, pudo finalmente elaborar una tabla con los logaritmos de los números enteros de 1 a 20.000 y de 90.000 a 100.000, que ha sido utilizada por científicos e ingenieros durante los últimos cuatro siglos para realizar toda clase de cálculos. Pueden encontrar en Internet la *Arithmetica* completa y anotada por Ian Bruce, de la Universidad de Adelaida. Los logaritmos también sirvieron para diseñar las *reglas de cálculo*, a las que INVESTIGACIÓN Y CIENCIA dedicó un interesante artículo el pasado mes de julio de 2006.

El trabajo de Napier y Briggs es realmente impresionante. Aún más si se tiene en cuenta que la notación decimal se había introducido en Europa apenas treinta años antes. (Napier fue uno de los primeros en utilizarla y propagarla.) A veces creemos que en aquella época el desarrollo científico era mucho más lento que en la actualidad. Sin embargo, el siglo XVII puede considerarse "frenético" desde el punto de vista científico. Mientras Napier y Briggs desarrollan los logaritmos, Kepler racionaliza la astronomía y Galileo describe el movimiento en términos matemáticos. Unos cincuenta años después del encuentro de Napier y Briggs, Newton concibe los *Principia*. Y en esta explosión intelectual, una de las más importantes para todo el pensamiento occidental, aquel encuentro en Edimburgo, en el verano de 1615, fue sin duda decisivo.

parr@seneca.fis.ucm.es

Juan M.R. Parrondo

La paradoja de San Petersburgo

En 1738, Daniel Bernoulli estudió un simple juego de azar sobre el que continúan reflexionando economistas, filósofos y matemáticos. Se comienza con un "bote" de dos euros. Se lanza una moneda al aire: si sale cruz, yo doblo la cantidad que hay en el bote; si sale cara, usted se lleva el bote disponible en ese momento. Es decir, si la primera tirada es cara, usted gana 2 euros, si la primera tirada es cruz y la segunda cara, gana 4 euros, si la primera cara sale en la tercera tirada gana 8 euros, y si la primera cara sale en la tirada n-ésima gana 2^n euros. Obviamente, lo que a usted más le conviene es que salga cara lo más tarde posible. En cualquier caso, usted gana siempre algo de dinero, por lo que es justo que yo le cobre alguna cantidad o *cuota* para permitirle participar en el juego. La pregunta que se hizo Bernouilli, y que en cierto modo sigue sin resolverse, es: ¿cuál es la cuota de entrada que se debería cobrar para que el juego fuera justo?

Para cualquier sorteo "normal", la cuota de entrada justa debe ser igual al valor medio de la ganancia. Por ejemplo, cuando compra un número en una lotería de, digamos, 100 números, si el premio es de 1000 euros, el precio del boleto, o cuota de entrada en el sorteo, debería ser de 10 euros. Para usted que compra el boleto, la probabilidad de ganar los 1000 euros es 1/100 y la probabilidad de no ganar nada es 99/100. Por tanto, el valor medio de la ganancia es $1000 \times (1/100) = 10$ euros. Por otra parte, el organizador de la lotería, con este premio y este precio por boleto, y siempre que venda todos los números, no gana ni pierde nada.

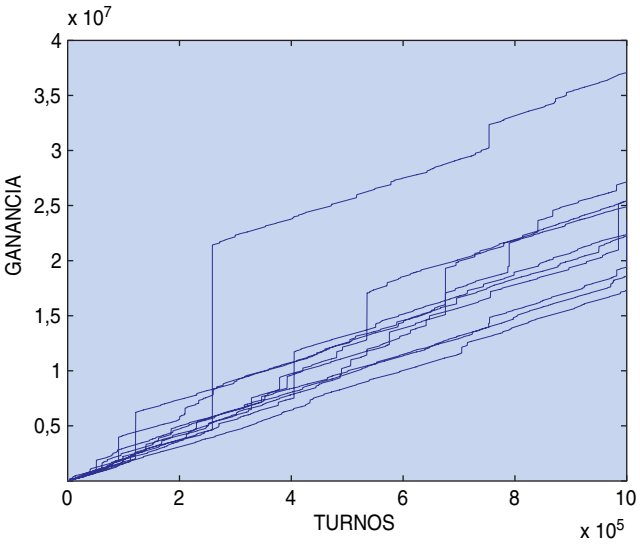
Si aplicamos este criterio al juego de San Petersburgo, nos encontramos con un serio problema. La probabilidad de que la primera cara salga en la tirada n-ésima es $(1/2)^n$, ya que, para que esto ocurra, debe salir cruz en las n-1 primeras tiradas y cara en la siguiente. En este caso la ganancia es 2^n. El valor medio de la ganancia es entonces:

$$G = \frac{2}{2} + \frac{4}{4} + \frac{8}{8} + \frac{16}{16} + \ldots$$

que es claramente infinito. Por tanto, la cuota de entrada debería ser infinita. En otras palabras, si yo le ofrezco entrar en el juego con una cuota de, digamos, un millón de euros, usted debería aceptar, porque la ganancia media en el juego, que es infinita, supera esa y cualquier otra cantidad. Sin embargo, nadie en su sano juicio aceptaría semejante trato. Esta es la paradoja de San Petersburgo: el sentido común nos dice que el valor medio de la ganancia no determina la cuota de entrada aceptable. ¿Cómo determinamos entonces dicha cuota?

El problema fundamental del juego de San Petersburgo es que proporciona premios muy cuantiosos con probabilidad extremadamente pequeña. Por ejemplo, si la primera cara aparece en la tirada décima, la ganancia es de 1024 euros, y esto ocurre con una probabilidad de 1 entre 1024. Las ganancias crecen exponencialmente mientras que las probabilidades decrecen también exponencialmente, siendo siempre el valor medio de cada posible premio igual a un euro.

En la figura 1 podemos ver diez "partidas" del juego de San Petersburgo, que constan de un millón de turnos del juego, que acaban cuando sale cara, con la ganancia correspondiente. He dibujado las ganancias acumuladas en función del número de turnos. La mayoría de las partidas acaban con una ganancia similar, en torno a los 20 millones, salvo una de ellas, en las que ha habido uno de esos eventos raros, en un turno un poco anterior al 300.000. En ese turno, salió cruz durante 23 tiradas seguidas y cara en la vigésimo cuarta, con lo que la ganancia del jugador fue de más de 16 millones de euros. La probabilidad de que esto ocurra es de una entre 16 millones. Sin embargo, en un millón de turnos, esto puede ocurrir con una probabilidad superior al 5 %. Por eso, no es sorprendente que en diez partidas de un millón de turnos cada una, hayamos sido testigos de un evento tan poco probable. En el resto de partidas también hay turnos con grandes ganancias: varios de ellos son saltos de unos 4 millones de euros y hay numerosos saltos de aproximadamente 2 millones y un millón de euros (recordemos que los saltos son siempre potencias de dos, pero 2^{20} es 1.048.576, número muy cercano al millón de euros). Cuantos más turnos juguemos, aparecerán saltos cada vez mayores. Eso es lo que hace



1. Diez partidas del juego de San Petersburgo, cada una de ellas de un millón de turnos.

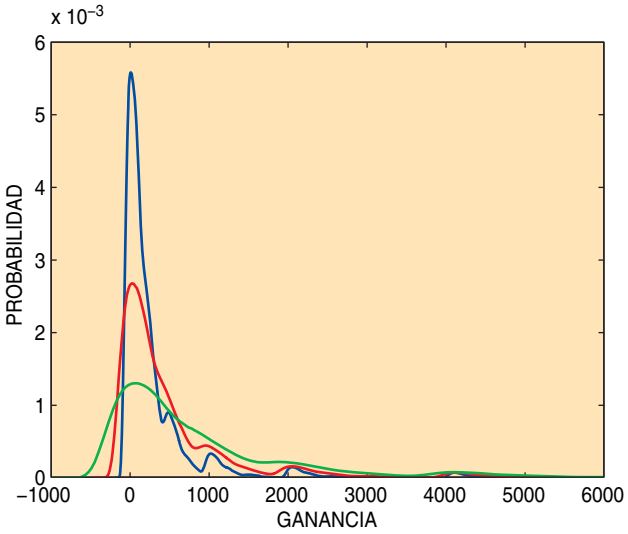
que el valor medio de la ganancia sea infinito. Pero si vamos a jugar sólo unas decenas de turnos, parece en principio absurdo que paguemos una cuota elevada por turno debido a eventos prácticamente imposibles. La paradoja sigue por tanto en pie.

Como decíamos al principio, la literatura sobre la paradoja de San Petersburgo es muy amplia y aún hoy en día se discuten posibles soluciones. Estos análisis y soluciones pueden clasificarse en dos grandes grupos: los que abordan el problema puramente matemático de un juego con ganancia media infinita y los que tratan de analizar cómo opera el "sentido común" y cómo valoramos el riesgo en un sorteo como el de San Petersburgo. Los trabajos que siguen esta última estrategia acaban casi siempre enfrentándose a cuestiones de índole psicológica. Uno y otro enfoque se corresponden con el problema al que se enfrenta el organizador y el jugador del sorteo, respectivamente.

Pongámonos en el papel del organizador de un sorteo de San Petersburgo. Somos dueños de un gran casino que quiere introducir este juego: ¿qué cuota de entrada deberíamos cobrar para hacer frente el pago de los premios? La diferencia entre el casino y el jugador es que el primero juega un gran número de veces y puede confiar en la estadística. Sin embargo, en el caso del juego de San Petersburgo, cuantos más turnos jugamos, más posibilidades hay de observar grandes saltos, como los de la figura 1. Varios matemáticos han estudiado la distribución de probabilidad de la ganancia del jugador después de un gran número N de turnos. Estas distribuciones están concentradas en torno a una cantidad que obviamente crece con N, pero que no es proporcional al número de turnos N, sino que vale $N \log_2 N$.

Por tanto, si el casino quiere recuperar el dinero que paga a los jugadores, tendrá que recaudar, después de N turnos, una cantidad $N \log_2 N$. De hecho, tendría que recaudar una cantidad algo mayor. En la figura 2 se muestra la distribución de probabilidad de las ganancias totales a las que he restado $N \log_2 N$, para N = 50, 100 y 200 turnos. Como vemos en la figura, las distribuciones están concentradas en torno a cero, como cabría esperar en cualquier sorteo justo, aunque la parte de ganancias positivas para el jugador es ligeramente superior a la de ganancias negativas. Podríamos corregir este sesgo aumentando ligeramente la cuota.

En cualquier caso, este resultado nos indica que la recaudación total no puede ser proporcional al número de turnos. Por el contrario, el casino debe cobrar *en cada turno* una cantidad ligeramente superior a $\log_2 N$, para así poder cubrir las pérdidas, siendo N el número *total* de turnos que está dispuesto a jugar con *todos* sus clientes. Esta solución a la paradoja, propuesta por Feller en 1968, puede que sea una



2. Distribución de probabilidad de la ganancia del jugador después de 50 (*curva azul*), 100 (*curva roja*) y 200 (*curva verde*) turnos, suponiendo una cuota total por participar $c = N \log_2 N$.

buena indicación de qué debe hacer el casino, pero, al cliente que quiere jugar al sorteo de San Petersburgo, le parecerá absurdo que le cobren por turnos en los que él no va a jugar. Incluso con un sólo cliente, una solución en donde la cuota dependa del número total de turnos es bastante peculiar. ¿Qué pasa si el casino y el cliente acuerdan prolongar la partida un millón de turnos más?

En 1985, el matemático sueco Martin Löf refinó el resultado de Feller, hallando la siguiente aproximación para la ganancia total G en N turnos:

$$\text{Prob}\left(\frac{G - N\log_2 N}{N} > x\right) \approx \frac{1}{x}$$

que es aceptable para valores de x superiores a 30. Esta fórmula se puede utilizar para encontrar cuotas más precisas que la dada por Feller. Por ejemplo, si el casino se contenta con que la probabilidad de perder dinero sea $1/x = 0{,}001$, tendrá que cobrar una cuota total $xN + N\log_2 N$, y una cuota por turno de $x + \log_2 N = 1000 + \log_2 N$ euros, que es superior a la sugerida por Feller en 1000 euros. Martin Löf señala que el término de Feller $\log_2 N$ es comparable con los 1000 euros sólo a partir de números N muy grandes. Por ejemplo, para un millón de turnos, dicho término es ligeramente inferior a 20. Por tanto, la cuota que se deduce de su análisis es prácticamente independiente de N.

Los resultados de Martin Löf son interesantes, pero no se pueden considerar una solución de la paradoja, puesto que nadie estaría dispuesto a pagar 1000 euros para participar en el juego de San Petersburgo. La razón es que la distribución de las ganancias tras un gran número de turnos es relevante para el casino, pero no para el jugador que sólo va a jugar unas pocas veces. En este caso, es difícil valorar los premios muy improbables y no hay un consenso acerca de cuál es la cuota de entrada que el jugador debería aceptar. El mes que viene analizaremos la paradoja desde esta perspectiva.

Juan M.R. Parrondo

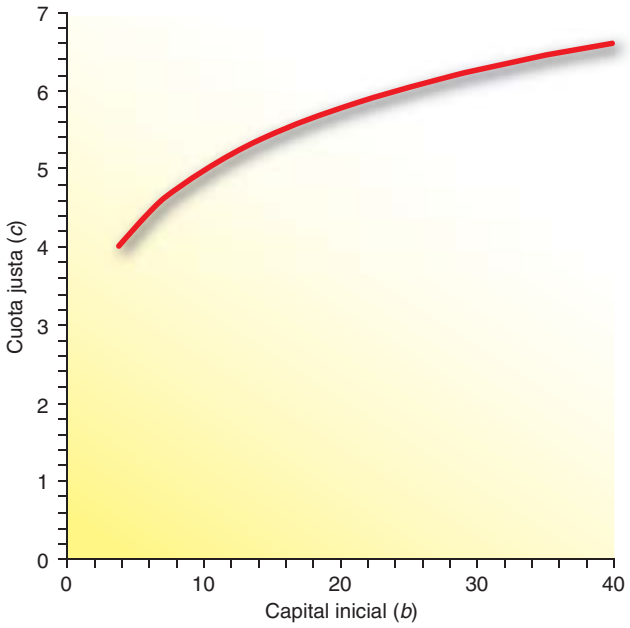
La paradoja de San Petersburgo y la teoría de la utilidad

El mes pasado analizamos la llamada *Paradoja de San Petersburgo*, un juego de azar en el que se lanza una moneda hasta que sale cara y el jugador gana 2^n euros si la primera cara ha salido en la tirada n-ésima. La paradoja surge cuando intentamos responder a la pregunta: ¿cuánto debería pagar un jugador para entrar en el juego? Normalmente, la cuota de entrada en un juego o sorteo debe ser igual a la ganancia media, pero, como vimos el mes pasado, en el juego de San Petersburgo esa ganancia media G_{med} es infinita. En efecto, la probabilidad de que la primera cara salga en la tirada n-ésima es $1/2^n$, de modo que:

$$G_{med} = \frac{2}{2} + \frac{4}{4} + \frac{8}{8} + \frac{16}{16} + \ldots$$

que es evidentemente infinita. De ello se deduce que uno debería entrar en el juego pagando cualquier cuota de entrada, por muy grande que ésta fuese, puesto que siempre la ganancia media será superior. Pero esta conclusión está en contradicción con el sentido común, ya que nadie en su sano juicio estaría dispuesto a pagar un millón de euros para participar en este sorteo.

Hace un mes analizamos la paradoja desde el punto de vista de un casino que pretende organizar el juego y tiene que decidir la cuota de entrada. Ahora vamos a estudiar la paradoja desde el punto de vista de un jugador que sólo va a jugar unas pocas veces. En este caso, la paradoja no tiene en realidad relación con el hecho de que la ganancia media sea infinita, sino con cómo valoramos el riesgo. La misma paradoja ocurre en el juego de San Petersburgo si limitamos los pagos, es decir, si acordamos que, tras por ejemplo 100 tiradas sin salir cara, el jugador no gana nada. En este caso, la ganancia media es de 100 euros; pero, ¿estaría alguien dispuesto a pagar esa cantidad para entrar en el sorteo? El premio puede ser aún enorme: $2^{100} \approx 10^{30}$ si la primera cara sale en la tirada número 100, aunque la probabilidad de ganar más de los 100 euros que cuesta entrar en el sorteo es bastante pequeña (1/64). Hay gente que podría considerar atractivo el sorteo con la cuota de 100 euros, a pesar de que lo más probable es que pierda dinero. Al fin y al cabo, muchos juegan a la lotería, y en ocasiones cantidades considerables, con la esperanza de ganar un premio muy cuantioso con una probabilidad insignificante. Sin embargo, hay una diferencia importante desde el punto de vista psicológico entre la paradoja y la lotería. En el juego de San Petersburgo, para ganar un premio cuantioso tiene que salir un gran número de cruces seguidas en el lanzamiento de la moneda, un evento muy improbable que puede ocurrir o no ocurrir aunque juguemos miles de veces. En el caso de la lotería, el boleto con el "gordo" *tiene que estar* en algún sitio y a *alguien* le tiene que tocar. Es muy distinto creer que ese alguien puedo ser yo a creer que, al lanzar una moneda, las 20 primeras tiradas van a ser cruz. Aunque esta diferencia en la percepción de uno y otro sorteo es irracional, porque ambos sucesos son casi igual de probables.



1. La cuota justa en función del capital inicial según Bernoulli.

Uno de los análisis más conocidos de la paradoja es el que Daniel Bernoulli publicó en la revista de la Academia de Ciencias de San Petersburgo en 1738. De hecho, la paradoja recibe su nombre precisamente por ese artículo. La fama del mismo no se debe a que ofreciera una solución completamente satisfactoria, sino porque en su trabajo Bernoulli sentó las bases de la *teoría de la utilidad*. La teoría trata de cuantificar el beneficio real o *utilidad* que supone para un individuo una determinada cantidad de dinero. El argumento fundamental de Bernoulli era razonable: 100 euros, por ejemplo, es una cantidad considerable para alguien que no posee nada, mientras que es insignificante para una persona que tenga un patrimonio de un millón de euros. Bernoulli propuso que el aumento de utilidad cuando una fortuna de x euros aumenta en una cantidad muy pequeña Δx debía ser $\Delta x/x$. Con este supuesto y un poco de matemáticas, se puede demostrar que el aumento de la utilidad cuando nuestra fortuna pasa de b euros a a euros es

$$\Delta U = \ln(a) - \ln(b) = \ln\left(\frac{a}{b}\right)$$

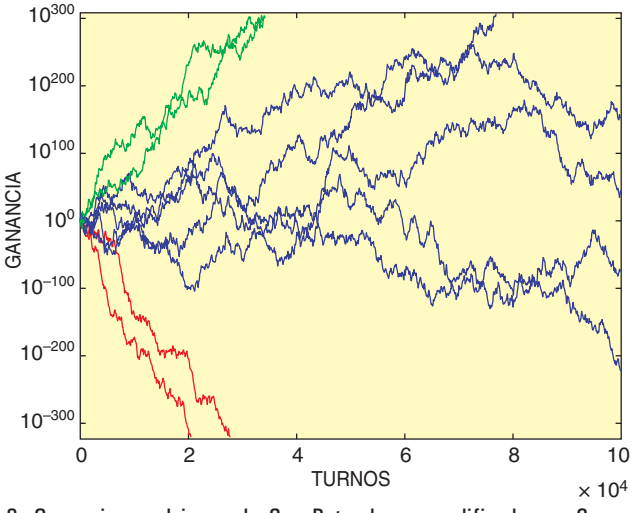
A partir de esta fórmula podemos calcular la utilidad media que uno obtiene si participa en el juego de San Petersburgo. Supongamos que la cuota de entrada es c y llamemos b a su capital inicial. El sorteo le puede deparar un premio de 2 euros con probabilidad 1/2. En este caso su nuevo capital sería de $b - c + 2$ euros.

Si gana 4 euros, lo que puede ocurrir con probabilidad 1/4, su nuevo capital será de $b - c + 4$ euros, y así sucesivamente. El incremento de utilidad que puede esperar, en media, del resultado del sorteo es:

$$\Delta U_{med} = \frac{\ln(b-c+2)}{2} + \frac{\ln(b-c+4)}{4} + \frac{\ln(b-c+8)}{8} + \ldots - \ln(b)$$

La cuota máxima que estaríamos dispuestos a pagar para entrar en el sorteo es la que hace que este incremento medio de la utilidad sea cero. Para hallar la cuota justa c, que será distinta según el capital inicial b, tenemos que solucionar la ecuación $\Delta U_{med} = 0$ con c como incógnita. En la figura 1 he dibujado la solución o cuota justa c en función del capital inicial b. La solución no se puede expresar en términos sencillos, aunque podemos calcularla para un caso particular, $b = c$, es decir, para el caso en que invertimos en el juego todo nuestro capital. ¿Cuándo merece la pena asumir este riesgo? Si hacemos $c = b$, nuestra ecuación $\Delta U_{med} = 0$ se simplifica y puede resolverse con algunos conocimientos de matemáticas. La solución es $c = 4$ euros. Si el capital inicial b es menor de cuatro euros, entonces la ecuación carece de solución, lo que indica que en este caso no merece la pena entrar en el juego por muy pequeña que sea la cuota. Por eso la curva de la figura 1 se interrumpe debajo de los 4 euros. Mas la cuota justa crece sin límite cuando lo hace el capital inicial. La solución no es del todo satisfactoria porque se basa en un supuesto acerca de cómo valoramos el dinero y porque la cuota depende del capital inicial. De todos modos, recordemos que tampoco era satisfactoria la solución del mes pasado, pues en ella la cuota depende del número total de turnos que se van a jugar.

Tom Cover, de la Universidad de Stanford, uno de los mayores expertos mundiales en Teoría de la Información y muy aficionado a las apuestas, analizó hace unos años la paradoja de San Petersburgo y ofreció una solución diferente e ingeniosa, que no hace referencia a ninguna valoración subjetiva del premio. Consideró una ligera y lógica modificación del juego original: podemos participar en el mismo pagando cualquier cantidad de entrada x, pero el premio recibido será proporcional a dicho pago.



2. Ganancia en el juego de San Petersburgo modificado por Cover con cuotas de 4 euros (*azul*), 3,9 euros (*verde*) y 4,1 euros (*rojo*).

Si el pago x es igual a la cuota c, entonces recibimos todo el premio; si es la mitad, recibimos la mitad del premio. En general, si el premio es P, recibiremos una cantidad xP/c. La segunda suposición, bastante más restrictiva, de Cover es que el jugador reinvierte todas sus ganancias en cada turno, de modo que, si su capital total inmediatamente antes del turno t es $X(t)$ y el premio en dicho turno es P_t, entonces:

$$X(t+1) = \frac{X(t)}{c} P_t$$

La evolución del capital recuerda la de una inversión continuada en Bolsa, como la que analizamos en los *Juegos Matemáticos* de septiembre de 2005. Se puede resolver tomando el logaritmo de la ecuación anterior:

$$\ln X(t+1) = \ln X(t) + \ln P_t - \ln c$$

y promediando:

$$\langle \ln X(t+1) \rangle - \langle \ln X(t) \rangle = \langle \ln P_t \rangle - \ln c$$

Como vemos, el logaritmo del capital aumenta en media si la media del logaritmo del premio es mayor que el logaritmo de la cuota. De forma similar a lo que ocurría en la solución de Bernoulli, la cantidad importante es el valor medio del logaritmo del premio y no el valor medio del premio (que es infinito). Con un cálculo similar al realizado en el análisis de Bernoulli, se obtiene $\langle \ln P_t \rangle = 2 \ln 2$. Por lo tanto, para que el logaritmo del capital no crezca ni disminuya en media (y no lo haga, por tanto, el capital), la cuota tiene que ser precisamente 4.

En la figura 1 podemos ver la evolución del capital en el juego de Cover para 100.000 turnos con distintos valores de la cuota. El capital tiene tantas fluctuaciones, que lo he dibujado en una escala logarítmica. He realizado cinco simulaciones del juego para una cuota igual a 4 euros, que pueden verse en azul en la figura, dos para cuota 3,9, en verde, y dos para cuota 4,1 en rojo. Para la cuota de 4 euros las fluctuaciones son enormes, positivas y negativas. Sin embargo, una ligera desviación de esa "cuota justa" hace que el capital crezca siempre a infinito (*curvas rojas*) o decrezca a cero (*curvas verdes*). La idea de Cover no es una solución de la paradoja, si bien constituye un método ingenioso para que el logaritmo aparezca sin utilizar valoraciones de los premios, que es en definitiva lo que hace la solución de Bernoulli.

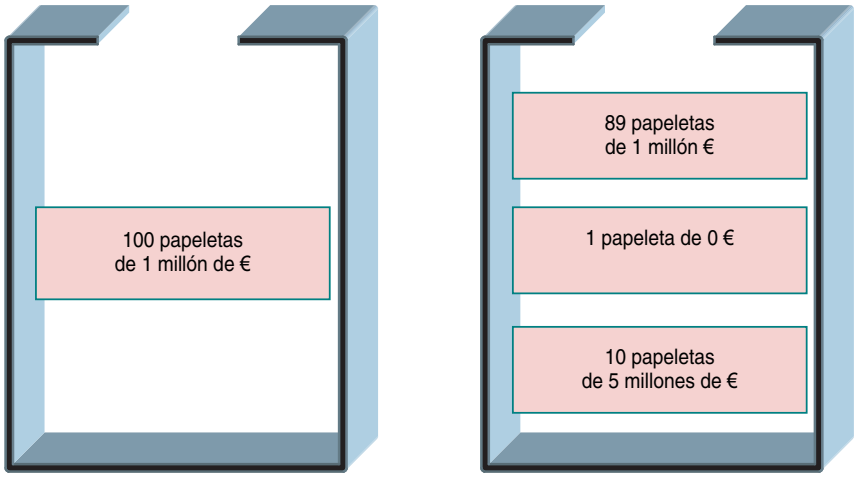
Hay algunos aspectos de la Paradoja de San Petersburgo que aún no hemos cubierto, como la solución de Euler, basada en medias geométricas de la ganancia (equivalente a la de Bernoulli). Sobre la paradoja han discutido grandes matemáticos, aportando observaciones cuando menos curiosas. A raíz del artículo del mes pasado, Fernando Pérez Dehesa me ha enviado unas notas, extraídas del libro de Mauricio Kraïtchik *Matemáticas Recreativas*, en las que se pueden leer comentarios muy interesantes de conocidos matemáticos del siglo XVIII y XIX. También quiero agradecer a otro lector, Carlos Macía, la sugerencia de discutir la paradoja en la sección. Como han podido comprobar, ha dado lugar a dos artículos y quizá nos depare más sorpresas en el futuro. ¡Anímense a sugerirme otros temas para la sección!

parr@seneca.fis.ucm.es

Juan M.R. Parrondo

Loterías y decisiones

Observe la figura 1. En ella hay dos urnas. La de la izquierda contiene 100 papeletas, todas y cada una de ellas con un premio de un millón de euros. La de la derecha contiene también 100 papeletas: 89 con un premio de un millón de euros, una papeleta sin ningún premio y 10 papeletas con un premio de 5 millones de euros. Usted tiene que elegir una de las dos urnas y sacar una papeleta de la urna seleccionada. El premio que contenga la papeleta será suyo ¿Cuál de las dos urnas elegiría? ¿Prefiere el millón de euros de la urna de la izquierda (1*a*) o se atreve con la de la derecha (1*b*), en la que puede ganar hasta 5 millones a costa de la posibilidad, poco probable, de irse a casa con las manos vacías?



1a *1b*

1. ¿Cuál de las dos urnas prefiere?

Observe ahora la figura 2. Tiene que elegir de nuevo entre dos urnas con 100 papeletas cada una. La de la izquierda, 2*a*, tiene 89 papeletas sin premio y 11 papeletas con un millón de euros. La de la derecha, 2*b*, tiene 90 papeletas sin premio y 10 papeletas con 5 millones de euros. ¿Cuál es en este caso su elección?

Prácticamente todo el mundo prefiere en el segundo caso la urna 2*b*. En el primero, es 1*a* la urna elegida mayoritariamente, aunque una minoría apreciable asume el pequeño riesgo de la papeleta sin premio y elige 1*b*. Si usted ha elegido 1*a* y 2*b*, entonces coincide con la mayoría de la gente, pero ha contravenido uno de los axiomas básicos de la teoría de la utilidad, una teoría que trata de cuantificar las preferencias de las personas ante situaciones inciertas.

El axioma que ha contravenido es el llamado axioma de independencia, que a primera vista parece bastante razonable. El axioma dice que, si prefiero una lotería *a* a una *b* y realizo sobre ambas una modificación idéntica, entonces la preferencia no debe cambiar, es decir, seguiré prefiriendo *a* sobre *b*. ¿Qué entendemos por una modificación idéntica? En nuestro ejemplo, significa añadir o quitar papeletas iguales de las dos urnas. En otras palabras, el axioma de indiferencia, en una formulación simplificada y adaptada a las urnas de las figuras, dice: si añadimos o retiramos de cada una de las dos urnas papeletas con el mismo premio, entonces el orden de preferencia no debe cambiar. ¿Les parece razonable? No debería ser así si han elegido 1*a* y 2*b*, porque esa elección está en franca contradicción con el axioma.

En efecto, tomen las dos urnas de la figura 1, retiren de las dos 89 papeletas con un millón de euros y reemplácenlas por 89 papeletas con 0 euros. El resultado son las dos urnas de la figura 2. Por lo tanto, si el axioma fuera cierto, usted debería haber mantenido la preferencia después del reemplazo de papeletas y haber elegido en las dos figuras las urnas de la izquierda o las de la derecha, pero nunca la 1*a* y la 2*b*.

Esta contradicción entre la elección de la mayoría de la gente y el axioma de indiferencia se llama paradoja de Allais, en honor del economista francés y premio Nobel Maurice Allais. No se trata en realidad de una paradoja, sino de una forma de probar que el axioma de independencia no reproduce adecuadamente nuestra valoración del riesgo.

Veamos en qué momento la preferencia entre las dos urnas cambia. Cuando retiramos las 89 papeletas con premio de un millón, nos quedamos con 11 de un millón en la urna de la izquierda y 1 de cero euros y 10 de 5 millones en la de la derecha. Creo que la mayor parte de la gente que elige 1.A mantendría su decisión tras este cambio. A continuación, añadimos, una a una, 89 papeletas sin premio. Tendremos primero 11 papeletas con un millón y una papeleta con cero euros en la izquierda y 10 papeletas con 5 millones y 2 con cero euros en la derecha. ¿Cambiaría entonces su preferencia hacia la urna de la derecha? ¿y después de añadir diez papeletas sin premio? ¿Puede precisar en qué momento cambia su decisión?

El axioma de independencia no es sólo razonable, sino que también es compatible con criterios matemáticos de decisión muy generales. Uno de estos criterios consiste en comparar los valores medios de la ganancia en cada una de las dos loterías. Este es de hecho el criterio adecuado si uno va a jugar un gran número de veces

al mismo sorteo. En la paradoja de Allais, el valor medio de la ganancia en la urna 1*a* es de un millón de euros; en la 1*b* es $(89\times1+10\times\times5)/100=1{,}39$ millones; en la 2*a* es $11\times1/100=0{,}11$ millones, y en la 2*b* el valor medio de la ganancia es $10\times\times5/100=0{,}5$ millones. Por lo tanto, el criterio del valor medio prescribe que 1*b* es mejor que 1*a* y que 2*b* es mejor que 2*a*.

Se puede demostrar que el criterio del valor medio verifica el axioma de independencia. De hecho, cualquier criterio basado en un valor medio lo verifica y es incompatible con la elección habitual en la paradoja: el par 1*a*, 2*b*. Supongamos que $U(p)$ es el beneficio o utilidad asociada a un premio p. Para que la elección sea el par 1*a*, 2*b*, tiene que verificarse:

$$100\times U(1)/100>1\times U(0)+89\times\times U(1)+10\times U(5)/100$$

$$89\times U(0)+11\times U(1)/\\/100>90\times U(0)+10\times U(5)/100$$

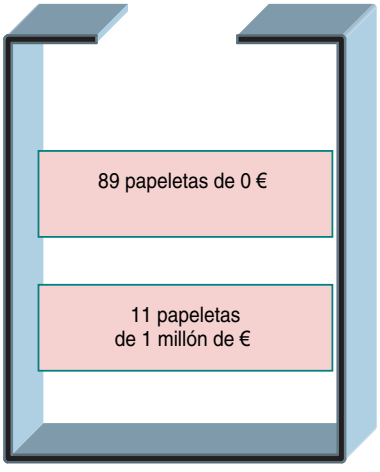
Pero estas dos desigualdades son incompatibles: si en la primera sumamos a ambos miembros $89\times U(0)/100$ y restamos $89\times U(1)$, obtenemos la segunda invertida. Estas sumas y restas no son más que la reproducción en lenguaje matemático de la operación de añadir o retirar papeletas de las urnas.

El mes pasado hablábamos de la paradoja de San Petersburgo y de cómo la solución que dio Bernoulli fue el comienzo de la teoría de la utilidad, a finales del siglo XVIII. En 1944, Von Neumann y Morgenstern desarrollaron una teoría axiomática en la que se incluía el axioma de independencia. La propuesta de Bernoulli que analizamos el mes pasado, en la que se utilizaba una función de utilidad $U(p)$ concreta, el logaritmo, también es incompatible con las preferencias habituales en la paradoja de Allais. Sin embargo, el propio Allais se considera un seguidor de Bernoulli, y a Von Neumann y Morgenstern, "neo-bernoullianos" que se han desviado de la idea original del científico suizo. Para Allais, esa idea original consistía en asignar un valor psicológico a cada premio, que puede depender no solo de su cuantía, sino también de su probabilidad (y que por supuesto puede ser distinto de un individuo a otro, dependiendo de su patrimonio, personalidad, etc.).

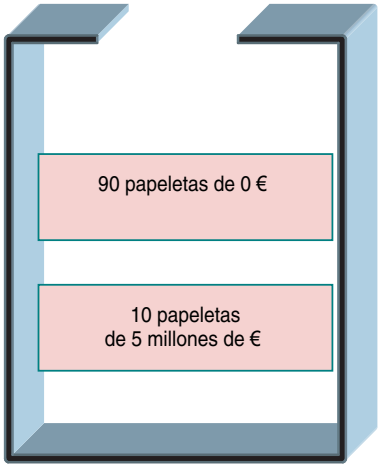
Paradoja de Ellsberg

Otra conocida paradoja relacionada es la de Ellsberg, en la que de nuevo tenemos que elegir entre varios sorteos. En una urna hay 90 bolas: 30 son rojas y el resto, 60, son negras o amarillas. Usted no sabe cuántas bolas negras y amarillas hay, tal y como se representa en la figura 3. Ahora tiene que elegir una de las siguientes opciones: 1.a) Si saca una bola roja le doy 100 euros; 1.b) Si saca una bola amarilla le doy 100 euros.

Tómese su tiempo. ¿Ha elegido ya? Si es así, considere esta nueva propuesta: 2.a) Si saca una bola roja o negra le doy 100 euros; 2.b) Si saca una bola amarilla o negra le doy 100 euros.

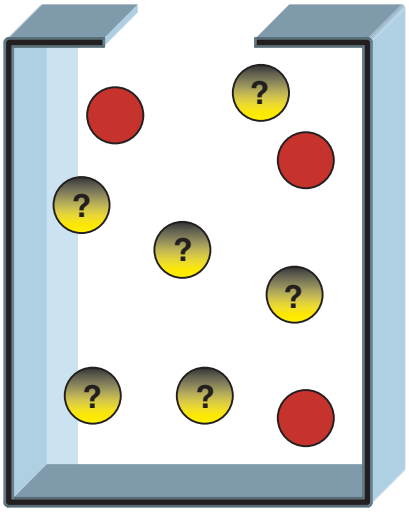


2a



2b

2. ¿Cuál de las dos urnas prefiere?



3. La paradoja de Ellsberg.

La mayor parte de la gente elige la opción 1.a) en el primer caso y la 2.b) en el segundo, pero esta elección contiene una contradicción. Eligiendo 1.a) estamos suponiendo implícitamente que hay más bolas rojas que amarillas, mientras que la elección de 2.b) significa que creemos que hay más amarillas que rojas. Podríamos pensar que la gente considera que 1.a) y 1.b) son prácticamente equivalentes, y que también lo son 2.a) y 2.b). Al fin y al cabo, lo serían si hubiera 30 bolas negras y 30 amarillas. Sin embargo, cuando se pregunta a los sujetos de estudio, no muestran indiferencia ante las dos opciones o una preferencia débil hacia una de ellas, sino que la preferencia es bastante intensa.

Lo que ocurre en la paradoja de Ellsberg es que la gente se ve atraída por la "tranquilidad" que proporciona un conocimiento completo de la situación. En el primer caso sabemos que, eligiendo a), tenemos una probabilidad del 30 % de ganar 100 euros, mientras que la opción b) es incierta y con ella no podemos ni siquiera evaluar las probabilidades de ganar. En el segundo caso pasa lo mismo: sabemos seguro que con la opción b) la probabilidad de ganar es del 60 %, mientras que no podemos evaluar esa probabilidad para la opción a). La paradoja de Ellsberg muestra otro rasgo interesante de nuestra relación con el riesgo: tratamos de evitar la ignorancia o, a la hora de evaluar sus consecuencias, tendemos a "ponernos en lo peor".

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

La asombrosa fórmula de Tupper

Uno de mis ilustres predecesores en esta sección fue Douglas Hofstadter. Asumió la difícil tarea de relevar a Martin Gardner en 1981 e imprimió a la sección un carácter peculiar, reflejo de sus propios intereses y obsesiones. De la lógica matemática y la inteligencia artificial a los grabados de Escher y las fugas de Bach, Hofstadter siempre ha reflexionado acerca de la naturaleza y los límites del pensamiento. En esta reflexión y en sus columnas de INVESTIGACIÓN Y CIENCIA, un tema recurrente fue la *autorreferencia*, es decir, la posibilidad de que una cierta entidad lingüística, matemática o computacional se refiera a sí misma.

La autorreferencia da lugar a razonamientos circulares y paradojas enigmáticas. Un ejemplo clásico es la paradoja del mentiroso, de la que existen múltiples versiones. La más simple consiste en una sola frase que afirma su propia falsedad: "Esta frase es falsa". No puede ser cierta ni falsa sin caer en una contradicción.

Consideren otro ejemplo lingüístico. Clasificamos los adjetivos calificativos en dos grupos: llamaremos *autóclitos* a los adjetivos que pueden aplicarse a sí mismos y *heteróclitos* a los que no pueden aplicarse a sí mismos. Por ejemplo, el adjetivo "esdrújulo" es autóclito, porque esdrújulo es una palabra esdrújula. Otro ejemplo es "pentasílabo" que es una palabra con cinco sílabas y, por tanto, pentasílaba. Es difícil dar más ejemplos (¿puede el lector encontrar alguno?), puesto que la mayoría de los adjetivos, como "blanco", "francés" o "estrecho", son heteróclitos. Ahora intente responder a la siguiente pregunta: ¿es el adjetivo "heteróclito" autóclito o heteróclito? Si es autóclito, entonces se aplica a sí mismo y es por tanto heteróclito; y si es heteróclito, entonces no se aplica a sí mismo y no puede ser heteróclito. De nuevo una frase: "el adjetivo 'heteróclito' es heteróclito", afirma su propia falsedad.

Para resolver, o al menos arrojar luz sobre este tipo de paradojas, los lógicos introdujeron la distinción entre un *lenguaje* y su correspondiente *metalenguaje*, que está formado por lo que se puede decir acerca del lenguaje original. En la autorreferencia una proposición dice algo de sí misma, y por tanto es al mismo tiempo lenguaje y metalenguaje.

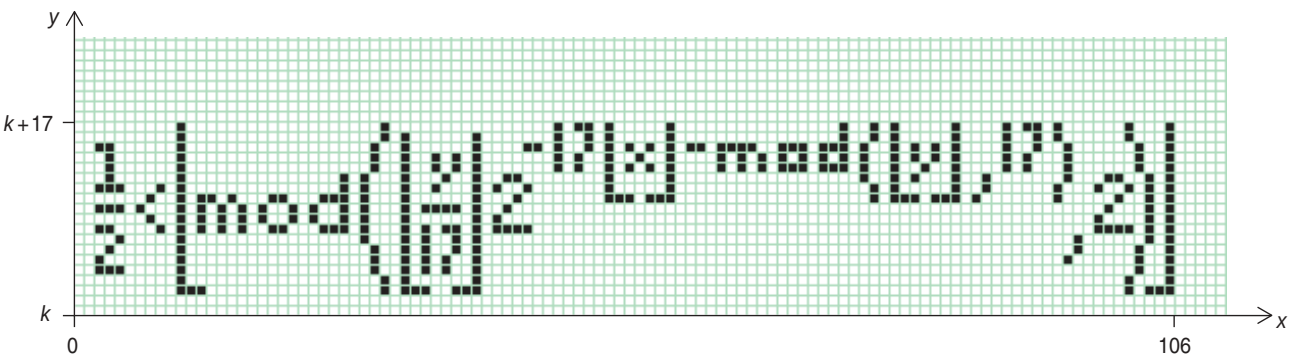
Kurt Gödel utilizó esta mezcla de lenguaje y metalenguaje no para construir paradojas, sino para demostrar uno de los resultados más profundos de la matemática, el *teorema de incompletitud*. El teorema afirma que cualquier sistema de axiomas que pretenda describir la aritmética de los números naturales es necesariamente incompleto, es decir, habrá afirmaciones verdaderas acerca de los números que no podrán demostrarse a partir de los axiomas.

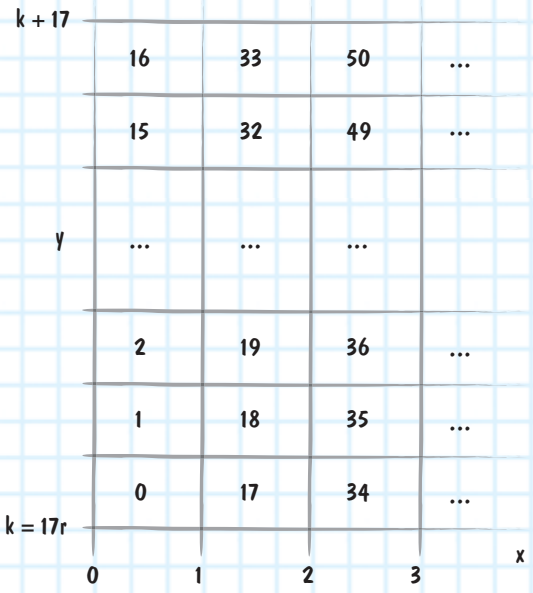
Gödel estableció una relación entre números y fórmulas, asignando a cada fórmula aritmética un número natural, y luego entre fórmulas y proposiciones metamatemáticas, es decir, proposiciones acerca de las propias fórmulas. Finalmente, construyó una fórmula G, que decía, en esa relación entre matemática y metamatemática: "yo no soy demostrable". Es decir, si G es verdadera entonces no es demostrable y si es falsa entonces es demostrable. Si el conjunto de axiomas es consistente, esta segunda posibilidad no puede darse (no se puede deducir algo falso de un sistema de axiomas consistente). Por lo tanto, G es indemostrable a partir de los axiomas.

El libro más conocido de Hofstadter lleva precisamente por título *Gödel, Escher, Bach* y en él trataba el teorema de Gödel y lo relacionaba con la inteligencia artificial, las fugas de Bach y los grabados de Escher. En sus artículos de *Investigación y Ciencia* escribió en numerosas ocasiones acerca de toda clase de autorreferencias, desde el teorema de Gödel hasta un curioso juego, el *Nomic*, en el que los jugadores cambiaban las reglas del propio juego a lo largo de la partida (de hecho, las reglas iniciales del juego simplemente establecían los mecanismos para modificarlas). Pero la autorreferencia no sólo da lugar a paradojas o a grandes teoremas. El más familiar y más misterioso fenómeno de la Naturaleza, la conciencia que los seres humanos tenemos de nosotros mismos, es en última instancia un tipo de autorreferencia.

Este mes recordamos al maestro Hofstadter con un nuevo ejemplo de autorreferencia, que sin duda habrá

1. El "autorretrato" de la fórmula de Tupper.





2. Cómo varía el exponente n en la fórmula de Tupper de píxel a píxel.

atraído su atención: la *fórmula de Tupper*, que tiene la sorprendente propiedad de dibujarse a sí misma en el plano. La fórmula es la siguiente:

$$\frac{1}{2} < \left\lfloor \mathrm{mod}\left(\left\lfloor \frac{y}{17} \right\rfloor 2^{-17\lfloor x \rfloor - \mathrm{mod}(\lfloor y \rfloor, 17)}, 2 \right) \right\rfloor$$

En ella aparecen dos símbolos que quizás algunos lectores no conozcan, aunque son habituales en matemáticas. El primero de ellos es $\lfloor x \rfloor$ y es la parte entera de x, es decir, el entero inmediatamente por debajo de x. Por ejemplo, $\lfloor 2{,}45 \rfloor = 2$. Como ven, la parte entera de un número es simplemente lo que hay a la izquierda de la coma decimal. La otra función que aparece en la fórmula de Tupper es la función "módulo", mod(x,z), que es igual al resto que se obtiene cuando se divide x entre y. Por ejemplo mod(36,17) = 2.

La fórmula establece una condición para el par x, y. Si coloreamos de negro en el plano (x,y) los puntos que verifican la condición, obtendremos un dibujo, una imagen infinitamente grande, puesto que el plano (x,y) es infinito. Pues bien, en un lugar de esa imagen infinita, en concreto, en la región definida por $0 < x < 106$, $k < y < k + 17$, en donde k es un número enorme, de 542 cifras, aparece lo que se muestra en la figura 1. ¿Les resulta familiar? ¡Es precisamente la propia fórmula de Tupper! Es decir, la fórmula se dibuja a sí misma, aunque en una remota zona del plano. Lo que están contemplando en la figura 1 es un auténtico "autorretrato" matemático.

¿Cómo llegó Tupper a concebir algo tan sorprendente y tan extraño? En realidad no es tan difícil, porque la fórmula de Tupper no sólo se dibuja a sí misma, sino que dibuja *cualquier* imagen de 17 píxeles de altura y ancho arbitrario. Para convencerse de ello hace falta un razonamiento un poco complicado, aunque sólo utiliza matemáticas elementales. Analicemos la fórmula cuando y varía desde k hasta $k + 17$, siendo k un múltiplo de 17, es decir, $k = 17r$, con r un número entero. En toda esta región del plano

$$\left\lfloor \frac{y}{17} \right\rfloor = r$$

es constante. Fijémonos ahora en el exponente de 2 en la fórmula (cambiado de signo), es decir, en la expresión:

$$n = 17\lfloor x \rfloor - \mathrm{mod}(\lfloor y \rfloor, 17)$$

En la figura 2 se muestra cómo este exponente n varía en el plano. Si dibujamos una trama de *píxeles* cuadrados de lado unidad, n toma un valor distinto en cada uno de ellos, tal y como se indica en la figura. Para convencerse de ello, basta darse cuenta de que mod($\lfloor y \rfloor$,17) es un entero que varía de 0 a 16 cuando y va desde $k = 17r$ hasta $k + 17$, mientras que el término $17\lfloor x \rfloor$ salta de 17 en 17 cuando avanzamos de una columna a la siguiente.

Veamos ahora al paso crucial para entender la fórmula de Tupper. La desigualdad

$$\frac{1}{2} < \left\lfloor \mathrm{mod}\,(r\,2^{-n}, 2) \right\rfloor$$

se puede analizar de forma sencilla si expresamos r en base 2. Supongamos que r es, en binario, 1011001 y que $n = 5$. Para dividir r entre 2^5, basta colocar la coma decimal cinco lugares a la izquierda de la última cifra, es decir, $r2^{-5} = 10{,}11001$. El módulo 2 de este número se obtiene eliminando las cifras a la izquierda de la primera cifra entera: mod($r2^{-5}$,2) = 0,11001. Finalmente, la parte entera de este número es simplemente la cifra que queda a la izquierda de la coma decimal, en este caso 0. Por lo tanto, el miembro de la derecha de la desigualdad no es más que la cifra $n + 1$ de r empezando por la derecha. Si esta cifra es un 0, la fórmula es falsa. Si es un 1, es verdadera y generará un píxel negro en el lugar correspondiente al exponente n.

Con todo ello podemos encontrar en qué región del plano dibuja la fórmula de Tupper cualquier imagen. Basta escribir un número binario entero r con unos y ceros en donde queremos que haya píxeles negros y blancos, respectivamente. Tenemos que escribir r de derecha a izquierda, mientras seguimos los píxeles de la imagen en el orden especificado en la figura 2 (desde $n = 0$ hasta el valor que queramos). Una vez que obtenemos r de esta forma, lo multiplicamos por 17 y obtenemos $k = 17r$. La imagen buscada estará en la franja que va de $y = k$ hasta $y = k + 17$ y de $x = 0$ hasta el valor de x en el que termina la imagen, que puede ser tan ancha como queramos (los píxeles a la derecha del último codificado en el número r serán blancos).

La fórmula no da lugar a ninguna paradoja, pero no es sólo autorreferente, sino una auténtica *Biblioteca de Babel*, de la que ya hemos hablado en alguna ocasión (*Caos, determinismo y voluntad*, julio de 2002 y *La paradoja de la Biblioteca de Babel*, octubre de 2003). Dibuja *todas* las posibles imágenes de 17 píxeles de altura y de anchura arbitraria, y como en 17 píxeles de altura es fácil dibujar cualquier letra del abecedario y cualquier signo de puntuación, entonces la fórmula "escribirá", en alguna remota franja del plano y como si se tratara de un rudimentario teletipo, el Quijote entero, desde su primera hasta su última letra, o cualquier otro libro escrito o por escribir.

Juan M.R. Parrondo

Pensamiento formal y pensamiento concreto

Observen la figura 1. En ella hay cuatro cartas. Cada una de ellas tiene en uno de sus lados una letra y en el reverso un número. Como ven, algunas presentan visible el número y otras la letra. Fíjense en la afirmación del recuadro que hay encima de las cartas: *Detrás de toda letra D está el número 3*. ¿Qué cartas tienen que voltear para comprobar que la afirmación es cierta? Una pista: la afirmación dice sólo que demuestren que detrás de una D hay un 3, pero no dice nada acerca del reverso de un 3. ¿Han resuelto ya el problema? Anoten la solución en un papel y pasemos a la siguiente prueba.

DETRAS DE TODA D HAY UN 3

1. ¿Qué cartas hay que levantar para comprobar la afirmación del cuadro superior?

Observen ahora la figura 2. Cuatro personas están en un bar consumiendo alguna bebida. En la figura se indica de cada uno de ellos o bien la edad o bien lo que están bebiendo. Usted tiene que comprobar que *todo el que bebe alcohol es mayor de 18 años*. ¿A quién tiene que preguntar la edad o lo que bebe para comprobar dicha afirmación?

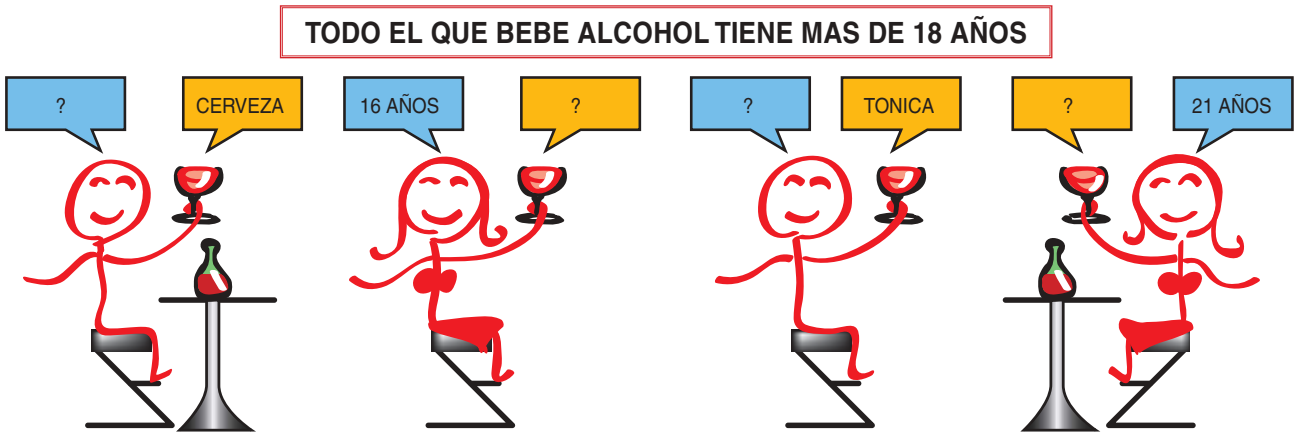
Supongo que este último "problema" les habrá resultado bastante trivial. La mayoría de la gente lo resuelve de forma casi inmediata: hay que preguntar la edad al individuo que bebe cerveza y la bebida que está consumiendo la chica de 16 años.

Sin embargo, en el problema de la figura 1 la mayoría de la gente se equivoca. Para confirmar la frase recuadrada hay que voltear la carta que tiene una D y comprobar que en efecto hay un 3 en su reverso, pero también la que tiene un 1, para comprobar que *no hay* una D en su reverso, puesto que en ese caso tendríamos una carta con el par D-1 y la afirmación recuadrada sería falsa. Si usted ha fallado en el problema de la figura 1, no se preocupe: sólo alrededor de un 10 % de la población encuentra la respuesta correcta y este porcentaje ni siquiera superó el 50 % cuando se realizó la prueba a profesores universitarios de matemáticas.

Lo curioso de este par de problemas es que formalmente son idénticos. En ambos se pide comprobar que cierta propiedad P (tener una D en una de sus caras en el primer caso, beber alcohol en el segundo) implica o está siempre acompañada de cierta propiedad Q (tener un 3 en una de las caras en el primer caso, ser mayor de 18 años en el segundo). En ambos se muestran cuatro casos: uno que verifica P, otro que no verifica Q, un tercero que no verifica P y un cuarto que verifica Q, tal y como se ve en la figura 3. El lector puede comprobar que incluso el orden en que están dispuestos las cartas de la figura 1 y los individuos de la figura 2 es *formalmente* el mismo que el de la figura 3.

De este pequeño experimento, diseñado por el psicólogo Peter C. Wason en 1968, se deduce que los seres humanos no pensamos formalmente, es decir, basándonos sólo en relaciones lógicas, sino que lo hacemos en función del *significado* de las afirmaciones que tenemos que demostrar o comprobar. Nuestro modo de pensar es semántico y no puramente lógico. A muchos puede parecerles esta afirmación evidente y un tanto trivial. Sin embargo, los que nos dedicamos a la matemática o a la enseñanza de la matemática tendemos a olvidarnos de ello. Son mayoría los matemáticos que piensan que la lógica formal es el esqueleto del pensamiento, sobre el que luego se añaden significados. Con esta concepción, muchos profesores de matemáticas nos desesperamos porque nuestros alumnos no son capaces de extrapolar un argumento de una situación a otra. Deberíamos recordar que el pensamiento formal es una *construcción* bastante elaborada y basada en el pensamiento referido a situaciones concretas, y que habría que partir de este último para llegar al primero. Creo que gran parte del fracaso escolar en matemáticas se debe a que los "espíritus geómetras", como decía Pascal, no somos conscientes de la artificialidad del pensamiento formal, que está basado en "principios [que] son palpables pero se apartan del uso común, de modo que nos cuesta volver la cabeza hacia ese lado, por falta de hábito".

El test de Wason es una herramienta excelente para investigar cómo razonamos. Los resultados del test ponen de manifiesto que utilizamos dos sistemas de razonamiento: uno inmediato, inconsciente e intuitivo, que en algunos artículos se denomina *sistema 1*, y otro formal



2. ¿A quién hay que preguntar la edad o lo que está bebiendo para comprobar la afirmación del cuadro superior?

y consciente, el *sistema 2*. Cuando nos presentan el test, el sistema 1 elige las cartas que aparecen en la afirmación que tenemos que comprobar, es decir, la carta con la D y la carta con el 3. Es lo que los psicólogos llaman "*matching bias*", que podríamos traducir como sesgo por afinidad. Después entra en juego el sistema 2, que corrige esta primera conclusión del sistema 1. En esta etapa, mucha gente se da cuenta de que no es necesario voltear la carta con el 3, puesto que sólo se nos pide comprobar que toda carta con una D tiene un 3 en su reverso, pero nada se informa de las cartas con un 3. En este análisis, sin embargo, la carta con el 1 pasa inadvertida para la mayoría de la gente (es lo que me pasó a mí cuando realicé el test). Como prueba de estos mecanismos, se ha realizado el test registrando el movimiento de los ojos de los participantes y se ha comprobado que éstos pasan la mayor parte del tiempo analizando las cartas D y 3.

Otra forma, más interesante y más extendida, de experimentar con el test de Wason es dotar de significado a las propiedades P y Q de la figura 3, tal y como hemos hecho con el ejemplo del bar de la figura 2. Existen numerosas variantes de este tipo clasificadas según el significado de P y Q.

El grupo de John Tooby, de la Universidad de California en Santa Bárbara, distingue tres situaciones diferentes: reglas descriptivas, reglas de contrato social y reglas de precaución. En la primera, P y Q son propiedades que sólo describen algún objeto o situación. El test original con las cartas es un ejemplo de este tipo de regla. Peor también los hay que utilizan propiedades del mundo real: comprueben la afirmación "siempre que se poda el jazmín a finales del invierno, florece en primavera", en los cuatro casos siguientes: 1) un jazmín que se ha podado, 2) otro que no ha florecido, 3) uno que no se ha podado y 4) otro que ha florecido. La respuesta correcta es: preguntar si 1) ha florecido y preguntar si 2) ha sido podado. ¿Encontraría el lector la respuesta correcta o pasarían inadvertidos los jazmines no florecidos?

El segundo tipo de situación es el contrato social, donde P consiste en algún tipo de beneficio y Q en alguna obligación. Por ejemplo, P puede ser "tomar prestado un coche" y Q "llenar el depósito de gasolina". En la tercera regla, peligro/precaución, P es la exposición a un peligro, como "trabajar con muestras contaminadas", y Q una medida de seguridad, como "utilizar guantes". Nuestro ejemplo del bar pertenecería a esta regla de precaución (aunque tiene también algo de contrato social). Tooby y sus colaboradores han llegado a la conclusión, mediante multitud de experimentos, que sólo entre un 20 % y un 30 % de la gente encuentra la solución correcta en las reglas descriptivas, mientras que el porcentaje sube al 70 % o el 80 % en las reglas de contrato social y las de precaución. De ello deducen que los seres humanos poseemos mecanismos de razonamiento intuitivo e inconsciente, adaptados a este tipo de situaciones y que se han desarrollado a lo largo de la evolución biológica porque eran especialmente útiles para la supervivencia de la especie.

Pero estas conclusiones se hallan sujetas a debate. Una explicación alternativa a los resultados del test de Wason es la *teoría de la relevancia cognitiva*, que afirma

3. ¿Qué información es necesaria para comprobar la afirmación del cuadro superior?

que nuestro sistema de razonamiento inconsciente otorga un peso a las distintas realizaciones del test de acuerdo con su relevancia cognitiva, es decir, con la capacidad de cada realización de cambiar nuestro conocimiento acerca del mundo. Este sistema de razonamiento es más general que la propuesta de Tooby, en la que se supone la existencia de sistemas de razonamiento específicos para cierto tipo de situaciones, como la regla de contrato social y la de peligro/precaución.

En cualquier caso, creo que la lección más importante que proporciona el test de Wason, especialmente para quienes enseñamos matemáticas, es que hay que tener cuidado cuando tratamos de desarrollar en nuestros alumnos el razonamiento formal. Por un lado, nos recuerda que el razonamiento formal es una elaboración artificial a la que no se tiene por qué acceder de un modo "natural". Por otro, el test nos puede ayudar a encontrar ejemplos eficaces para enseñar ciertas reglas lógicas.

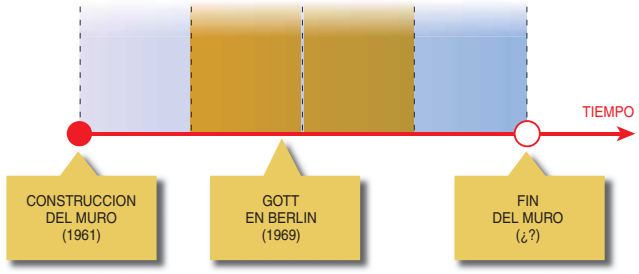
JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Sutilezas estadísticas

En los años noventa circuló en ambientes científicos un sorprendente resultado estadístico que más tarde resultó ser un fiasco: la *regla de Gott*, un método sencillo, pero sorprendentemente general, para estimar la duración de un fenómeno cualquiera.

J. Richard Gott III es un imaginativo cosmólogo de la Universidad de Princeton que ha trabajado en temas muy sugerentes, como la posibilidad de viajar en el tiempo. En 1969 realizó un viaje por toda Europa y, ante el Muro de Berlín, se le ocurrió el siguiente argumento. Su visita a Berlín no tenía nada de especial. Podría haber sucedido en cualquier momento de la existencia del Muro. Cabe pensar entonces que ocurre en un instante al azar de la vida del Muro. Si dividimos dicha vida en cuatro partes iguales, tal y como se muestra en la figura 1, la probabilidad de que la visita de Gott ocurra en uno de esos cuatro intervalos es 1/4. Por tanto, la probabilidad de que la visita se dé en los dos intervalos centrales (*marrón en la figura 1*) es 1/2. Pero, si la visita acontece en estos dos intervalos, el Muro no puede durar menos de 8 + 8/3 años, ni más de 8 + 3 × 8 años.



1. El argumento *copernicano* de Gott: la probabilidad de que la visita de Gott al Muro de Berlín se realizara en el intervalo de tiempo indicado por los rectángulos rayados es 1/2.

La primera cota se obtiene suponiendo que la visita se produce al final de los dos intervalos centrales, con lo que la duración de cada intervalo sería de 8/3 y al Muro le quedaría sólo un intervalo, es decir, 8/3 de año, y su duración total sería la suma de los cuatro intervalos, 8/3 × 4 años. La segunda cota se obtiene suponiendo que la visita se produce al principio de los dos intervalos centrales, con lo que cada intervalo dura 8 años y al muro le quedan esos dos intervalos más el cuarto, es decir, tres intervalos de 8 años de duración. Por lo tanto, con probabilidad 1/2, la vida del Muro tiene que estar comprendida entre 10 años y 8 meses y 32 años. Es decir, el Muro, razonó Gott en 1969, caerá, con probabilidad 1/2, entre 1972 y 1993.

Veinte años después de su visita a Berlín, en 1989, el muro caía; Gott recordó su viejo argumento y pensó que merecía la pena ahondar algo más. El argumento puede aplicarse para estimar la duración de cualquier fenómeno, sin más que saber cuándo comenzó y suponiendo que lo observamos en un instante completamente aleatorio dentro del tiempo de vida del mismo.

A esta suposición Gott la llama *principio copernicano*. Copérnico fue quien propuso una nueva imagen del cosmos en la que la Tierra ya no era el centro del universo, como creía el sistema ptolemaico. En el universo copernicano la Tierra no tiene un estatuto especial, sino que ocupa una región más del universo. Esta declaración, extendida también al tiempo, se conoce en cosmología como principio copernicano. Según el principio copernicano ni la Tierra ocupa un lugar especial en el universo ni el presente tiene nada de especial en la historia del mismo. Para Gott, la expresión matemática de ese principio, aplicada a la observación de cualquier fenómeno independiente del observador, es que el tiempo t de observación es un instante elegido al azar dentro del intervalo de vida del fenómeno.

El argumento de Gott puede generalizarse a márgenes de confianza superiores a 1/2. Supongamos que un fenómeno se inició en un tiempo t_0 y que lo observamos en un tiempo t. Si la duración total del fenómeno es T, supondremos que t es un número aleatorio entre t_0 y $t_0 + T$. Entonces, la probabilidad de que t se encuentre entre $t_0 + aT$ y $t_0 + (1 - a)T$ es la fracción que ocupa el área rayada en la figura 2, es decir:

$$\mathrm{Prob}\,[t_0 + aT \le t \le t_0 + (1-a)T] = 1 - 2a$$

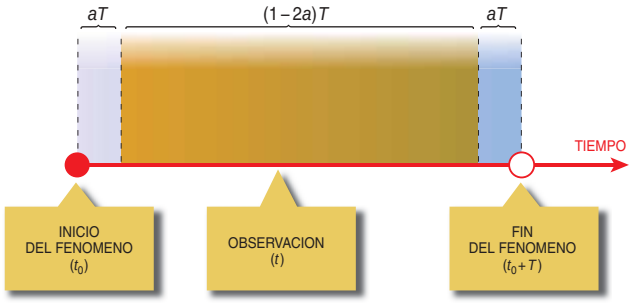
Consideramos ahora los dos casos extremos: que la observación ocurre al principio del intervalo rayado, $t = t_0 + aT$, o que ocurre al final de dicho intervalo, $t = t_0 + (1 - a)T$; y obtenemos así la cota superior e inferior para la duración T del fenómeno:

$$\frac{t - t_0}{1 - a} \le T \le \frac{t - t_0}{a}$$

Esta desigualdad se debe satisfacer con una probabilidad $1 - 2a$, que suele denominarse *nivel de confianza*. Por ejemplo, para $a = 1/4$, obtenemos un nivel de confianza del 50 % y el intervalo que dedujo Gott en su visita a Berlín en 1969. Si queremos un nivel de confianza del 95 %, tendremos que utilizar $a = 0{,}0025$, con lo cual el intervalo de confianza es más amplio:

$$\frac{40}{39}(t - t_0) \le T \le 40\,(t - t_0)$$

En su artículo en *Nature* de 1994, Gott aplicó esta fórmula a toda clase de fenómenos y situaciones: la vida de la Unión Soviética, la del mismo Gott, la duración de la especie humana o los años en que seguirá publicándose la propia revista *Nature*. Por ejemplo, la especie humana se originó hace unos 200.000 años. Si aplicamos el principio copernicano a nuestra época,

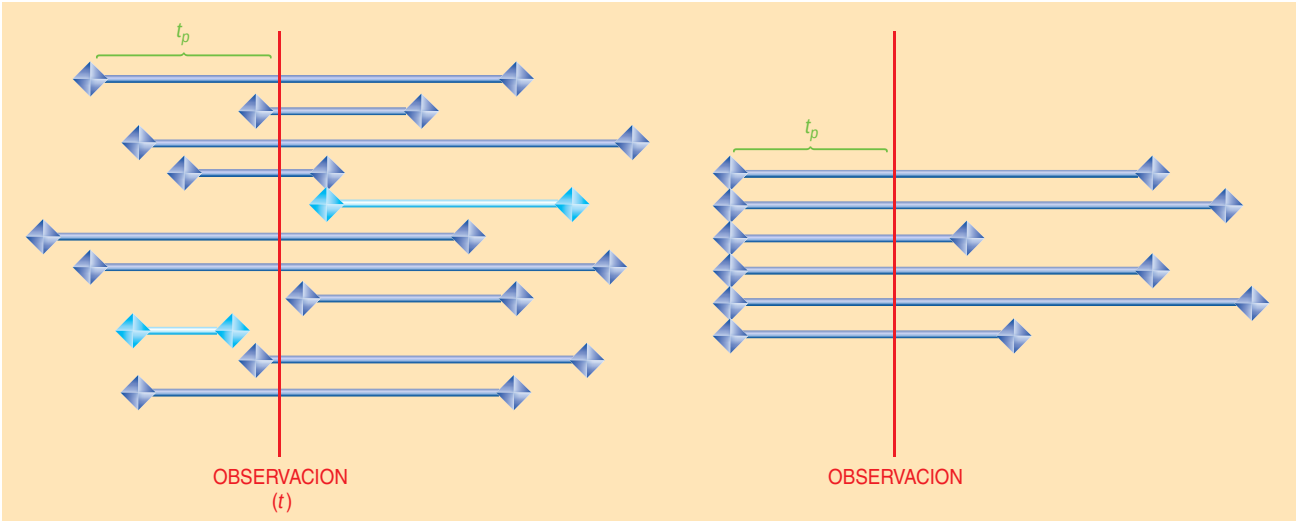


2. Generalización del argumento de Gott.

considerándola un instante cualquier en la historia de la especie y aplicamos el argumento de Gott, obtenemos que, con probabilidad del 95 %, la humanidad sobrevivirá al menos 5100 años pero no más de 7,8 millones de años. Aplicado a una vida humana, el argumento da resultados aparentemente triviales o absurdos. Yo tengo 43 años, por lo que, según Gott, viviré entre 44,1 y 1720 años. Pero un bebé de un mes vivirá, con una probabilidad del 95 %, entre 40/39 = 1,026 y 40 meses, es decir, poco más de 3 años. Del argumento de Gott parece deducirse que el 95 % de los niños de un mes no van a vivir más de 40 meses, lo cual es evidentemente falso. El argumento también falla si lo aplicamos al fenómeno "no salir el cero en la ruleta", puesto que el número de turnos que lleva sin salir el cero no tiene la más mínima influencia sobre las probabilidades de que salga o no salga el cero en los turnos siguientes. En estos casos, ¿no es aplicable el principio copernicano o el argumento es simplemente incorrecto?

Carlton Caves publicó en el año 2000 un artículo muy crítico contra el argumento de Gott, en el que realizaba un análisis *bayesiano* del mismo. Ya hemos hablado en esta sección de este tipo de análisis (*Fósiles y loterías,* abril 2005). El error del argumento de Gott es muy sutil, pero puede entenderse con un sencillo ejemplo. La mejor forma de reflexionar sobre argumentos probabilísticos es suponer que podemos repetir un experimento un gran número de veces. En este caso, admitamos que tenemos muchas copias del fenómeno cuya duración queremos estimar. Tal y como se muestra en la figura 3 (*izquierda*), el comienzo de estas copias y su duración son cantidades aleatorias. De acuerdo con el principio copernicano, en cierto instante al azar t intentamos observar el fenómeno. Lo lograremos en unas ocasiones y en otras no, como se indica en la figura. Vamos a analizar sólo las ocasiones en las que observamos el fenómeno. El tiempo t_p transcurrido entre el comienzo del mismo y la observación es una cantidad variable. El argumento de Gott se basa en que t_p/T es un número aleatorio uniformemente distribuido entre 0 y 1. Esto es rigurosamente cierto. Para convencerse de ello, imaginen la figura 3 (*izquierda*) con un ingente número de copias y piensen en la cantidad t_p/T, que es la fracción de segmento que queda a la izquierda de la observación en cada copia. Dicha cantidad puede tomar cualquier valor entre 0 y 1 con igual probabilidad.

Sin embargo, si ahora de entre todas esas copias nos quedamos con las que poseen un determinado valor para t_p, según se muestra en el esquema de la derecha de la figura 3, esta hipótesis deja de ser cierta. Imaginen que la duración T es un número aleatorio entre 1 y 10 años y que t_p es 6 años. Entonces, las copias que han sobrevivido a nuestra criba, es decir, las que aparecen en la parte derecha de la figura 3, durarán 6, 7, 8, 9 o 10 años, con probabilidad 1/5. Por lo tanto, t_p/T puede ser igual a 1, 6/7, 6/8, 6/9 o 6/10 y toma cada uno de estos valores con probabilidad 1/5. Como vemos, t_p/T está muy lejos de ser un número uniformemente distribuido entre 0 y 1. De hecho, lo es sólo cuando la probabilidad de la duración T tiene una forma muy particular (proporcional a $1/T^2$). La conclusión de Caves es muy sutil: la hipótesis de Gott es verdadera *sólo* si no conocemos t_p; pero si no conocemos t_p, ignoramos cuánto tiempo lleva ocurriendo el fenómeno desde su comienzo hasta la observación y, por tanto, no podemos usar el argumento. Por otro lado, si conocemos t_p, la hipótesis es falsa. En pocas palabras, el argumento de Gott o bien es falso o bien es inútil.



3. Si no fijamos el tiempo t_p transcurrido entre el comienzo del fenómeno y la observación (*esquema de la izquierda*), la hipótesis de Gott (t_p/T es una cantidad aleatoria entre 0 y 1) es válida. Sin embargo, si fijamos t_p (*esquema de la derecha*) la hipótesis deja de ser cierta.

JUEGOS MATEMÁTICOS

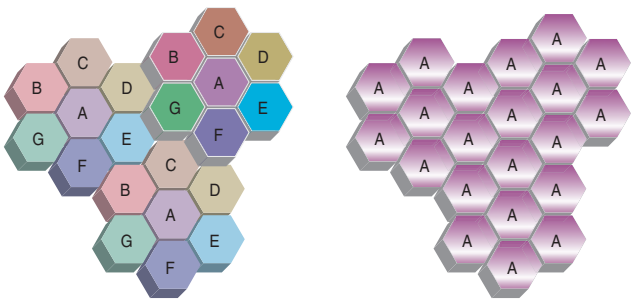
Juan M.R. Parrondo

Móviles y vectores

Termina el primer tiempo del partido en un gran estadio. En ese momento, miles de personas toman su teléfono móvil para llamar a casa o a algún amigo. Aunque en este tipo de situaciones las líneas suelen saturarse, son muchos los que consiguen llamar. ¿No se han preguntado nunca cómo es posible que la información proveniente de todos esos aparatitos pueda procesarse por separado? ¿Y cómo unas pocas antenas pueden distribuir a su vez información a todos ellos, de forma dirigida, específica y bastante rápida? Denle las gracias a las matemáticas. Más concretamente, a los vectores y matrices que tantos quebraderos de cabeza provocan en nuestros estudiantes de secundaria y de universidad.

En los comienzos de la telefonía móvil, la estrategia para que la estación base se pudiera comunicar con varios teléfonos al mismo tiempo fue asignar a cada uno de ellos una frecuencia o canal, como si se trataran de distintas cadenas radiofónicas. Este sistema de acceso múltiple por división de frecuencia (FDMA, acrónimo de Frequency Division Multiple Access) imperó en la telefonía móvil hasta los años noventa. Pero no utiliza el ancho de banda disponible de manera óptima. Entre otras razones porque, para evitar interferencias, cada antena o base no puede emplear todas las frecuencias disponibles. Las antenas de telefonía móvil se disponen formando una red aproximadamente hexagonal, como la que se muestra en la figura 1. Debido a posibles interferencias, una celda no puede utilizar las mismas frecuencias que sus vecinas, lo cual obliga a una cuidadosa planificación de las frecuencias asignadas a cada celda, como vemos en la figura 1 (*derecha*). Las celdas con distintas letras emplearán frecuencias diferentes; así, cada una de ellas hará uso sólo de un séptimo del ancho de banda disponible, con la pérdida consiguiente de capacidad para toda la red.

La segunda generación de telefonía móvil se basó en el sistema de acceso múltiple por división de tiempo, TDMA (de Time Division Multiple Access). En este sistema no sólo se divide el espectro de frecuencias, sino también el tiempo. En cada canal habla más de un usuario (3 en EE.UU. en canales de 30 kHz, y 8 en Europa y Asia en canales de 200 kHz) en intervalos de tiempo muy cortos, del orden de 30 o 40 milisegundos. Cada usuario utiliza este intervalo de tiempo de forma secuencial; el oído apenas nota las breves interrupciones, aunque hay una pérdida de calidad apreciable. Es un hecho reconocido que la segunda generación de telefonía móvil perseguía el aumento de capacidad, aunque fuera a costa de la calidad de la comunicación. Sin embargo, en este sistema, sigue siendo necesaria la planificación de frecuencias y continúa sin aprovecharse todo el ancho de banda disponible.



1. Planificación de celdas y frecuencias en FDMA y TDMA (*izquierda*) y en CDMA (*derecha*).

Por último, la 3G o tercera generación de telefonía utiliza el sistema de acceso múltiple por división de código, CDMA (de Code Division Multiple Access). Permite aprovechar, de manera muy ingeniosa, la misma frecuencia (con un ancho de banda mayor) para *todos* los usuarios que se encuentran en el área de alcance de una determinada antena o base. No se requiere ya ninguna planificación de frecuencias, como se muestra en la figura 1 (*izquierda*); cada celda de la red puede servirse, en principio, de todo el ancho de banda disponible.

El CDMA ha sido posible gracias a la digitalización de la señal que viaja de nuestros teléfonos a la base o antena más cercana (aunque esta digitalización ya se encontraba en el TDMA). El teléfono convierte nuestra voz en una señal digital compuesta por ceros y unos. La base también transmite al teléfono una señal digital y éste la convierte en el voltaje que finalmente reproduce el sonido en el altavoz. Sin embargo, para explicar el funcionamiento del CDMA es mejor recurrir a los valores +1 y –1, en vez de unos y ceros. En esta representación, que se denomina polar, +1 equivale a un uno y –1 equivale a un cero.

A cada teléfono que recibe o realiza una llamada en el área correspondiente a una base, ésta le asigna un número binario de varios dígitos, código que podemos representar como una serie de números, en este caso +1 o –1 (en matemáticas, un *vector*). Un código de cinco dígitos podría ser, por ejemplo:

$$\mathbf{b} = (+1, -1, -1, +1, -1)$$

El teléfono necesita enviar y recibir constantemente bits, es decir, ceros y unos. Sin embargo, en el sistema CDMA no los envía directamente, sino que transmite el código **b** si el bit que quiere mandar es uno y –**b** si el bit que quiere mandar es cero. Utilizando la representación polar con +1 y –1, podemos expresar tal prescripción de la siguiente forma matemática: si lo que el teléfono quiere enviar es *s*, deberá enviar *s***b**. Como el código **b** tiene determinada longitud, 5 dígitos en nuestro ejemplo, aunque 64 o más en las ejecuciones reales del CDMA,

necesitamos enviar todos estos bits para transmitir el bit *s*. Tamaño "derroche" nos permite identificar la fuente del bit entre todos los usuarios que se están comunicando con la base en un momento dado.

En la figura 2 podemos ver cómo funciona la codificación CDMA con el código **b** de cinco dígitos que hemos puesto de ejemplo. Según se aprecia en la figura, el número de bits que tiene que enviar el teléfono se multiplica por cinco. Normalmente un teléfono necesita enviar o recibir 9600 bits por segundo, para ello envía o recibe 1.228.800 bits por segundo utilizando distintos códigos, algunos propios del teléfono y otros propios de la base o asignados por ESTA a una determinada conversación.

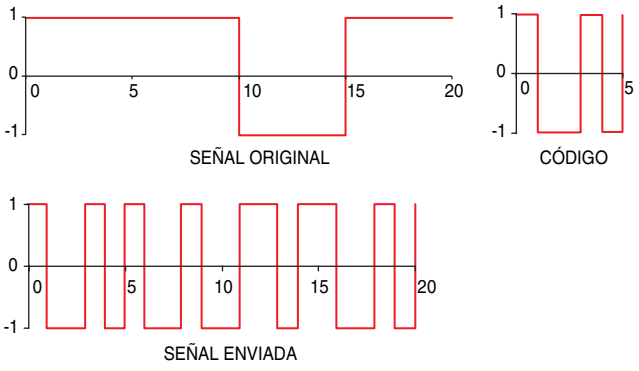
Cada teléfono o usuario tiene su propio código. Pese a ello, todos envían sus señales en la misma frecuencia. Por lo tanto, la base recibe la suma de todas ellas. Si están hablando simultáneamente 5 usuarios, por ejemplo, y si sus envíos están sincronizados, la señal que recibe la base en un intervalo de tiempo de longitud igual a la longitud de los códigos, es:

$$\mathbf{s} = s_1\mathbf{b}_1 + s_2\mathbf{b}_2 + s_3\mathbf{b}_3 + s_4\mathbf{b}_4 + s_5\mathbf{b}_5$$

en donde s_1, ..., s_5 son los bits que cada usuario quiere transmitir y $\mathbf{b}_1$, ..., $\mathbf{b}_5$ los correspondientes códigos. ¿Cómo puede la base inferir de esta suma la contribución de cada usuario? E este punto, el álgebra de vectores puede ayudarnos.

Una operación básica con vectores es el llamado *producto escalar*, que consiste en multiplicar las componentes de dos vectores una a una y sumar todos estos productos. Por ejemplo, el producto escalar de **a** = (1,5,4) y **b** = (3,0,2) es **a**·**b** = 1 × 3 + 5 × 0 + 4 × × 2 = 11. Cuando el producto de dos vectores es nulo, decimos que son *ortogonales*. La denominación proviene de que, en espacios de dos y tres dimensiones, el producto escalar de dos vectores que forman un ángulo de 90 grados es nulo.

Sin embargo, la ortogonalidad es una propiedad más general. Así, en el caso de vectores cuyas componentes son sólo +1 y –1, que el producto escalar sea nulo es equivalente a que difieran en la mitad de sus componentes. Los vectores ortogonales nos van a permitir resolver el problema del CDMA. Supongamos que los



2. La codificación de la señal en el sistema CDMA. La señal original se multiplica por el código del usuario y se envía a la base el resultado.

códigos asignados a cada usuario son ortogonales entre sí; en tal caso, si multiplicamos escalarmente la señal recibida **s** por uno de los códigos, por ejemplo $\mathbf{b}_1$, el producto de dicho código por todos los demás será nulo. Tenemos también que

$$\mathbf{b}_1.s = s_1\mathbf{b}_1.\mathbf{b}_1 = Ns_1$$

siendo *N* la longitud del código, ya que al multiplicar cada componente de $\mathbf{b}_1$ por sí misma el resultado es siempre 1 y la suma de todos esos productos conduce a $\mathbf{b}_1 \cdot \mathbf{b}_1 = N$. Por lo tanto, si los códigos son ortogonales, la base puede recuperar la información enviada por cada usuario simplemente multiplicando escalarmente la señal recibida por el código correspondiente. El sistema de envío de información de la base a los teléfonos es similar. En este caso la base envía la suma **s** y son los teléfonos los que extraen de ella la información dirigida a cada uno.

La pregunta crucial para la capacidad de la red es entonces ésta: ¿cuántos códigos ortogonales existen y cómo podemos obtenerlos? El CDMA es ingenioso, pero no mágico: el número de códigos ortogonales es igual a la longitud de los mismos, o, formulado de modo más intuitivo y geométrico, el número máximo de vectores ortogonales en *N* dimensiones es precisamente *N*. Una forma particular de obtenerlos es a partir de las llamadas matrices de Hadamard-Walsh. Se comienza con la matriz:

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

y se genera una nueva matriz mediante el siguiente algoritmo:

$$H_2 = \begin{pmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Continuando de forma similar, se obtienen matrices en donde todas las columnas son vectores ortogonales entre sí. En la telefonía actual se utilizan estos códigos de Walsh con una longitud de 64 bits (aunque se extenderán a 256 en el próximo estándar de CDMA).

El sistema CDMA completo es bastante más complicado de lo que hemos esbozado aquí. Los códigos de Walsh se aplican a la transmisión base-teléfonos; para el envío de información de los teléfonos a la base se recurre a otro tipo de códigos que son sólo aproximadamente ortogonales. Si lanzamos dos monedas muchas veces, en más o menos la mitad de las tiradas ambas presentarán el mismo resultado. Por ello, dos vectores formados por +1 y –1 aleatorios tienden a ser ortogonales, lo cual permite construir códigos aproximadamente ortogonales mediante algoritmos muy simples que generan bits pseudoaleatorios. Probablemente hablemos de estos algoritmos en los próximos meses.

Juan M.R. Parrondo

Números pseudoaleatorios

Aunque pueda parecer extraño, la aleatoriedad es un bien preciado. Ejércitos y empresas de criptografía pagan considerables cantidades de dinero por una fuente segura de números aleatorios, que son también esenciales en una gran variedad de simulaciones por ordenador: fluidos, meteorología, materiales, moléculas, redes sociales, etc.

¿Por qué resulta valioso algo tan ubicuo como el azar? ¿No basta con tirar una moneda al aire para conseguir un resultado aleatorio? En efecto, pero en las aplicaciones mencionadas se necesitan grandes cantidades de bits aleatorios producidos de modo automático. Hay además un cierto grado de "calidad" en las fuentes de números aleatorios. Por ejemplo, el viejo lenguaje de programación BASIC disponía de una función que generaba números aleatorios utilizando lo que el reloj del propio ordenador marcaba cuando la función se ejecutaba.

Es cierto que el instante en que la función se ejecuta depende de muchas variables: momento en que se inicia el programa, longitud de éste, etcétera. Resulta prácticamente impredecible. Sin embargo, las llamadas consecutivas a esa función de números aleatorios no garantizan que éstos gocen de una distribución uniforme y carezcan de correlación alguna. Lo mismo ocurre con los lanzamientos de una moneda, que puede estar ligeramente sesgada.

Pero no faltan algoritmos deterministas que producen números con una aleatoriedad falsa, aunque de "mayor calidad". Como el algoritmo que los genera es predecible, se denominan pseudoaleatorios. Sin embargo, poseen muchas de las propiedades de los números aleatorios, las suficientes como para que sean idóneos en aplicaciones científicas y de simulación. En criptografía se utilizan también; mas, si se necesita un alto grado de seguridad, se recurre a otros métodos de generación, basados en el azar real de fenómenos físicos, como el ruido en los dispositivos eléctricos o incluso las desintegraciones radiactivas.

Uno de los métodos habituales para obtener números pseudoaleatorios son los registros de desplazamiento. Se trata de una serie de registros en donde se almacenan bits, es decir, ceros o unos, que se van desplazando consecutivamente y de modo circular. Si tenemos, por ejemplo, siete registros con los valores iniciales 1011100, desplazamos todos hacia la derecha y el último de ellos, el 0, lo colocamos al principio, el l resultado será 0101110. Después del segundo desplazamiento obtenemos 0010111 y así sucesivamente. Obviamente, el estado inicial se recupera tras siete pasos, dando lugar a una secuencia bastante trivial. Sin embargo, una sencilla modificación de estos desplazamientos circulares genera secuencias bastante complejas. La modificación consiste en colocar como primer dígito no el último registro, sino una combinación de éste y algunos otros, que se llaman llaves. El algoritmo resultante se denomina registro de desplazamiento con retroalimentación lineal o LFSR (por Linear Feedback Shift Register).

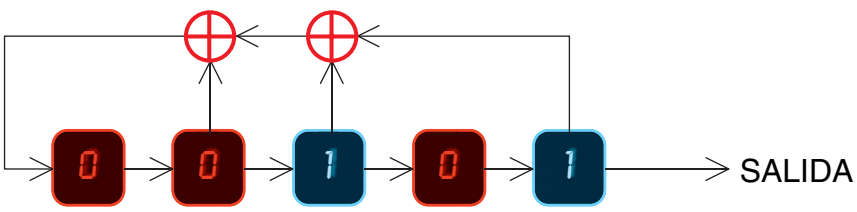
En la figura 1 podemos ver un LFSR con cinco registros. En cada desplazamiento, el primero se obtiene combinando tres llaves: el último, el segundo y el tercer registro. Estas combinaciones se representan en la figura mediante flechas y el símbolo $\oplus$ que denota una operación lógica muy simple entre dos números binarios: si los dos números son iguales, el resultado de la operación será 0; si son distintos, el resultado será 1. La operación se conoce como "disyunción exclusiva" o *exclusive or* (XOR) en inglés, porque, si 1 representa "cierto" y 0 "falso", entonces p $\oplus$ q es equivalente a "p o q, pero no las dos a la vez". Matemáticamente la operación $\oplus$ tiene también una interpretación sencilla. Es la suma módulo 2 de sus dos operandos:

$$x \oplus y = (x + y) \bmod 2$$

en donde *a* mod 2 es igual al resto cuando se divide *a* entre 2, es decir, 0 si a es par y 1 si es impar.

En el LFSR, el primer registro es simplemente la suma módulo 2 de todas las llaves. Para la configuración de la figura 1, las tres llaves tienen valores 0, 1 y 1 respectivamente, por lo que 0 $\oplus$ 1 $\oplus$ 1 = 0 es el nuevo valor del primer registro. Desplazamos los demás hacia la derecha y obtenemos la combinación 00010. Un nuevo paso daría 00001, y así sucesivamente. De todas estas configuraciones, vamos tomando como salida del algoritmo en cada paso el último de los registros. La cadena de bits de salida en nuestro ejemplo sería entonces 10100... Los cinco primeros bits de esta cadena son los de la configuración inicial. Mas, a partir del sexto, la secuencia adquiere mayor complejidad.

Como vemos, un LFSR queda determinado por el número de registros y por las llaves utilizadas. Como el último de los registros es siempre una llave (si así no ocurriera, el LFSR sería equivalente a uno con menos registros), basta indicar todas las llaves para describir un LFSR. Las llaves suelen escribirse en orden decreciente entre corchetes, de modo que la primera da el tamaño total



1. Esquema de un registro de desplazamiento con retroalimentación lineal (LFSR).

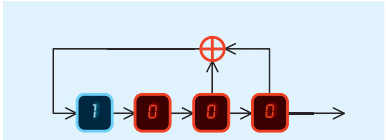
del LFSR. El de la figura 1 sería el [5 3 2].

Una familia de LFSR especialmente útiles para la generación de números pseudoaleatorios la observamos en los maximales. Un LFSR maximal de n registros pasa por todos y cada uno de los posibles estados, salvo por el compuesto por n ceros. Como hay 2^n posibles estados con n registros, este tipo de LFSR da lugar en su salida a series de bits que se repiten sólo cada $2^n - 1$ pasos. Basta con 20 registros para obtener un período de un millón de bits. Con 50 registros, el período de la secuencia alcanza tal magnitud, que cualquier aplicación práctica queda muy lejos del mismo.

Pero no todos los LFSR son maximales. Veamos un par de ejemplos con cuatro registros: el [4 3] y el [4 3 2]. En las tablas se distingue la evolución de los estados en cada uno de los registros, empezando ambos con 1000. He coloreado en rojo los dígitos llave con los que se obtiene, mediante suma módulo dos, el primer dígito en el siguiente paso. El primer LSFR es maximal, generando una secuencia de período 15. Sin embargo el segundo, a pesar de tener más llaves, no lo es. El estado original se recupera al cabo de 7 pasos. Por lo tanto, la secuencia de salida en este LFSR se repite cada 7 bits.

Cuando un LFSR es maximal no sólo genera una cadena de bits de período largo (el más largo posible), sino que estas cadenas guardan un estrecho parecido con cadenas de bits genuinamente aleatorios. En todas ellas hay prácticamente la misma cantidad de unos que de ceros (el número de unos es igual al número de ceros más uno). Unos y ceros que se hallan distribuidos de manera muy parecida a como los estarían si fueran aleatorios. Consideremos las subcadenas de dígitos consecutivos iguales y sus tamaños. Por ejemplo, en la cadena generada por [4 3], 000100110101111, las secuencias consecutivas tienen longitudes 3(0), 1(1), 2(0), 2(1), 1(0), 1(1), 1(0), 4(1); es decir, hay cuatro secuencias de longitud 1 (dos de ceros y dos de unos), dos secuencias de longitud 2 (una de ceros y una de unos), una secuencia con 3 ceros y otra con 4 unos.

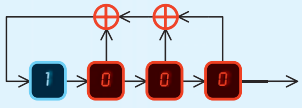
La salida de cualquier LFSR maximal tiene ese tipo de propiedad: hay un total de 2^{n-1} secuencias consecutivas, de las cuales la mitad presenta longitud 1, una cuarta parte longitud 2, 1/8 longitud 3, etc. En cada caso, siempre la mitad de ellas son de unos y la otra mitad de ceros, salvo para secuencias consecutivas cercanas al número de registros: sólo hay una con $n - 1$ ceros y otra con n unos. Todo esto es precisamente lo que se esperaría, aunque con fluctuaciones, en una larga cadena de bits genuinamente aleatorios.



PASO	REGISTROS	SALIDA
0	1000	0
1	0100	0
2	0010	0
3	1001	1
4	1100	0
5	0110	0
6	1011	1
7	0101	1
8	1010	0
9	1101	1
10	1110	0
11	1111	1
12	0111	1
13	0011	1
14	0001	1
15	1000	0

Los LFSR maximales son, por tanto, ideales como generadores de números aleatorios. De hecho, el principal objetivo de la teoría matemática de los LFSR es caracterizar cuáles de ellos son maximales. La caracterización completa requiere una matemática ligeramente refinada, basada en la factorización de polinomios. Pero algunas características de los LFSR maximales son sencillas: deben tener un número par de llaves (incluido el último registro) y las llaves deben ser números primos entre sí.

Las cadenas generadas por LFSR maximales guardan aún más similitudes con cadenas aleatorias. Sin embargo, como es evidente, son cadenas completamente deterministas: conocido el LFSR y su estado inicial, podemos predecir la cadena entera. ¿Hay algún modo de distinguir entre una de estas cadenas y una genuinamente aleatoria? Los LFSR maximales suelen pasar casi todas las pruebas de aleatoriedad cuando se aplican a cadenas mucho menores que el período.



PASO	REGISTROS	SALIDA
0	1000	0
1	0100	0
2	1010	0
3	1101	1
4	0110	0
5	0011	1
6	0001	1
7	1000	0

Existen algoritmos diseñados específicamente para detectar LFSR. De hecho, una definición de complejidad, la complejidad lineal, se basa en los LFSR: la complejidad lineal de una cadena de bits es el mínimo LFSR que la genera. Hay un algoritmo para encontrar la complejidad lineal de una cadena y, por tanto, el LFSR que la genera. Este algoritmo forma parte de la batería de pruebas de aleatoriedad estándar.

En cualquier caso, la aleatoriedad sigue siendo un fenómeno misterioso. La única definición objetiva y basada e en las propiedades de una cadena y no en el método como se ha obtenido es la dada por Kolmogorov. La complejidad de Kolmogorov de una cadena es la longitud del programa más pequeño que la genera. La complejidad lineal constituye, en realidad, un caso particular, pero sólo contempla programas basados en LFSR, mientras que la de Kolmogorov hace referencia a cualquier programa capaz de correr en cualquier ordenador. Una cadena aleatoria es aquella cuya complejidad es igual a su tamaño, es decir, es incompresible. Pero no existe ningún método para calcular la complejidad de Kolmogorov. Para eso tendríamos que ser capaces de detectar cualquier pauta o diseño, por ocultos que estuvieran, en los ceros y unos de una cadena de bits.

JUEGOS MATEMÁTICOS

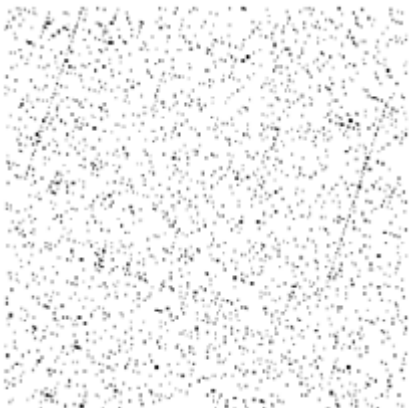
Juan M.R. Parrondo

Más sobre números aleatorios

Le propongo un reto aparentemente sencillo: ¿podría "imitar" el comportamiento de una moneda o de un dado? Chris Wetzel, del Rhodes College, en los EE.UU., creó una curiosa página web (http://www.rhodes.edu/Psych/wetzel/random/intro.html), en la que se puede comprobar lo difícil que es comportarse de forma puramente aleatoria. Se introduce en un casillero una serie "imaginaria" de tiradas, que no es más que una secuencia de H y T (por cara, *head*, y cruz, *tail*). Se tienen que introducir 100 tiradas, para lo cual no se tarda más de un minuto, puesto que sólo hay que teclear las dos letras mencionadas de la forma más aleatoria posible.

La serie entonces se analiza mediante diversos *tests de aleatoriedad*. Una serie realmente aleatoria puede consistir en cualquier secuencia de resultados, pero si la cadena es muy larga debe poseer, con una alta probabilidad, ciertas propiedades. Por ejemplo, el número de caras y el número de cruces deberían ser similares. En una serie de 100 tiradas, estos números deberían estar entre las 40 y las 60 tiradas. Pero hay otras propiedades menos evidentes. [ESO VALE SI LA PROBABILIDAD DE CARA Y CRUZ ES ½. SI FUERA 0,9999 CARA, LA ALEATORIEDAD CONSISTIRÍA EN LA IMPREDECIBILIDAD DE LA APARICIÓN DE LAS RARAS CRUCES EN UNA SERIE SUFICIENTEMENTE LARGA. SE PODRÍA HACER TAMBIEN LO DEL PROGRAMA DE WETZEL PARA ESE CASO].

En la página de Wetzel se realizan seis comprobaciones: 1) el número de caras; 2) el número de subsecuencias con el mismo resultado consecutivo; 3) el número mayor de caras o cruces consecutivos; 4) la probabilidad de que salga una cara habiendo salido una cara en el turno anterior; 5) la probabilidad de que salga una cara habiendo salido una cruz en el turno anterior; y 6) la diferencia entre las dos probabilidades anteriores. Con algunos conocimientos de probabilidad, se pueden calcular los valores medios de estas cantidades y el intervalo en que se encontrarán con un cierto nivel de confianza, por ejemplo, del 99 %. No les revelaré estos valores para que puedan hacer la prueba sin ningún conocimiento previo. Si la secuencia que usted ha tecleado da lugar a valores que están fuera de estos intervalos, habrá sido "descubierto" (aunque un 1 % de secuencias realmente aleatorias también fallará el test).



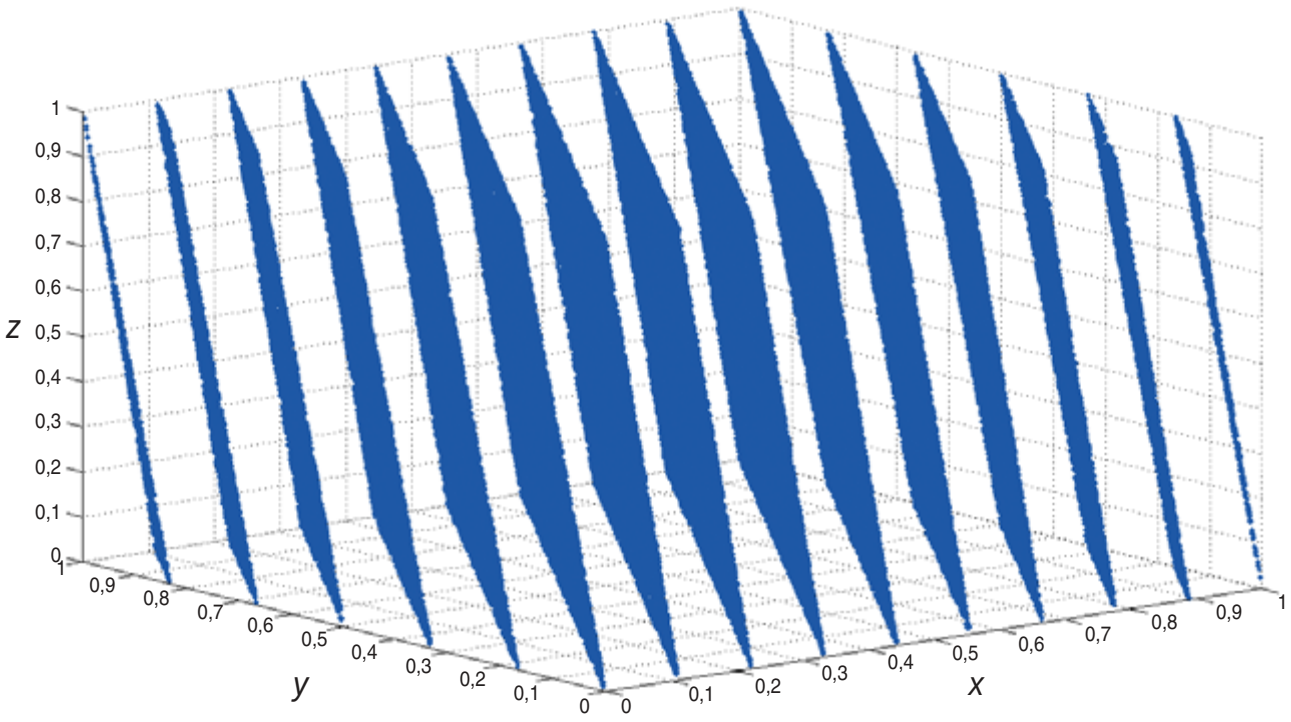
1. 5000 puntos en dos dimensiones generados por un algoritmo de congruencia lineal, con a = 1277, b = 131.072 y semilla I_n = 1, con 5000 puntos. Las dos líneas oblicuas no aparecerían en una secuencia realmente aleatoria.

Con un poco de práctica, uno puede pasar todos estos tests, aunque cualquier pequeño detalle se traduce en alguna pauta que es detectada. Por ejemplo, es bastante difícil teclear una buena secuencia aleatoria con las dos manos o incluso si se teclea H y T con dos dedos de una misma mano. En cualquiera de estos casos, surgen algunas asimetrías entre la H y la T. Incluso aunque el número de resultados H y de T' sea similar, los tres últimos tests, probabilidades de que a una cara le siga una cara o una cruz, son muy difíciles de superar.

Hay un sinfín de tests de aleatoriedad más complicados que los que utiliza Wetzel. ¿Por qué tanto interés en comprobar la aleatoriedad de una secuencia de números?

Como ya analizamos el mes pasado, los números aleatorios son necesarios en muchas situaciones. La más obvia es la organización de un sorteo o una lotería. En criptografía cumplen un papel crucial, al servir de plantilla para codificar mensajes. En el mes de agosto vimos también que los protocolos de comunicación entre nuestros teléfonos móviles y las centralitas pueden ser más eficientes si disponemos de números aleatorios. Finalmente, en muchas simulaciones numéricas de todo tipo, reacciones químicas, fluidos, mercados financieros, climatología o reactores nucleares, es necesario utilizar una gran cantidad de números aleatorios, que en ocasiones sirven para reflejar la aleatoriedad ineludible de lo simulado y en otras para explorar de forma eficiente el conjunto de estados del sistema o los posibles valores de los parámetros que lo definen.

La aleatoriedad real en una serie de números, es decir, su absoluta impredecibilidad y la ausencia de cualquier pauta en los mismos, sólo es posible si esos números se obtienen de algún proceso natural. Se puede por supuesto lanzar una moneda y anotar el resultado, pero en la mayoría de las aplicaciones se necesita una gran cantidad de números aleatorios, que sólo pueden generarse mediante métodos automatizados. Hay algunas empresas que ofrecen listados de números genuinamente aleatorios, obtenidos a partir de fenómenos cuánticos, como las desintegraciones radiactivas, de ruido en dispositivos electrónicos o de datos climatológicos, como la temperatura o la humedad en un cierto lugar. En la página web random.org se ofrecen series de números aleatorios basados en este tipo de datos y hay también información interesante sobre la generación de números aleatorios.



2. Diagrama tridimensional del IBM RANDU con 100.000 puntos (tomado de Wikipedia).

Ni siquiera con estos métodos automatizados es posible generar la cantidad de números aleatorios necesaria para muchas simulaciones. Se recurre entonces a números *pseudoaleatorios*, como los que vimos el mes pasado. Son secuencias de números obtenidos mediante algoritmos completamente deterministas. Por tanto, las secuencias son predecibles. Sin embargo, su comportamiento es muy parecido al de los números genuinamente aleatorios, y son capaces de pasar muchos de los tests de aleatoriedad. Algunos de ellos son además muy rápidos desde el punto de vista computacional, lo que los hace idóneos para muchas aplicaciones.

El mes pasado vimos un tipo de algoritmo que genera números pseudoaleatorios: los basados en desplazamientos de registros. Los teléfonos móviles utilizan constantemente uno de ellos para identificarse ante la base. Pero los generadores más extendidos, sobre todo en simulaciones numéricas, son los de *congruencia lineal*. La versión más simple consiste en la fórmula siguiente:

$$I_{n+1} = (a \times I_n) \bmod b$$

En esta fórmula a y b son números que caracterizan el algoritmo. La operación módulo, x mod b consiste en dividir x entre b y tomar el resto. I_n forma la secuencia de números aleatorios, que debido a la operación módulo, estarán entre 1 y $b - 1$ (el cero no es un valor deseable, porque, si en la secuencia aparece un cero, todos los valores siguientes serán también nulos). Si se desean números entre 0 y 1, basta dividir los I_n entre $b - 1$. Para comenzar la secuencia se escoge un valor inicial I_1, que suele llamarse *semilla*. Como la fórmula es completamente determinista, en cuanto vuelve a aparecer la semilla en la secuencia, los valores siguientes volverán a ser los mismos. En otras palabras, la secuencia es periódica y su período no puede ser mayor que $b - 1$. Lo interesante es entonces que b sea muy grande y que la secuencia tenga período máximo. Pero ni siquiera esto garantiza que la secuencia genere unos "buenos" números aleatorios.

Esta falta de calidad es en ocasiones difícil de detectar. Por ejemplo, con $a = 1277$, $b = 131.072$ y semilla $I_n = 1$, se obtiene una secuencia de período 32.769. No es de período máximo pero, hasta los 32 mil primeros números, pasa sin dificultad muchos de los tests de aleatoriedad. Sin embargo, si uno dibuja en el plano los pares de puntos (I_1,I_2), (I_3,I_4), etc., se encuentra con la figura 1, donde se han representado 5000 puntos. Las dos líneas oblicuas que aparecen en la gráfica indican una pauta que no estaría presente si la secuencia fuera genuinamente aleatoria.

Este tipo de representación gráfica de los números generados por el algoritmo es un buen test de aleatoriedad. Una representación similar en tres dimensiones de puntos con coordenadas (I_1,I_2,I_3), (I_4,I_5,I_6), ..., "desenmascaró" un algoritmo ampliamente usado en los años sesenta del siglo pasado para todo tipo de simulaciones, ya que estaba detrás de la función RANDU de los ordenadores de IBM. El RANDU utiliza $a = 65539 = = 2^{16} + 3$ y $b = 2^{31}$ y no muestra ninguna pauta en una representación bidimensional. Lo que se ve al hacerla es una nube de puntos uniformemente distribuida sobre el plano. Sin embargo, la representación tridimensional mostró lo que se puede contemplar en la figura 2. Todos los puntos se encuentran en 15 planos paralelos. De hecho, esta correlación tan significativa se puede demostrar de forma relativamente sencilla y es ahora sorprendente que nadie reparara en ella durante casi una década.

Semejante "fiasco" puso en duda muchos de los resultados de las simulaciones numéricas realizadas en los años sesenta y setenta y dio un toque de atención a los científicos. Pocos se fían ya de los generadores de números aleatorios que proporcionan los sistemas operativos o los lenguajes de programación comerciales y los diseñan ellos mismos o los obtienen de centros de investigación fiables.

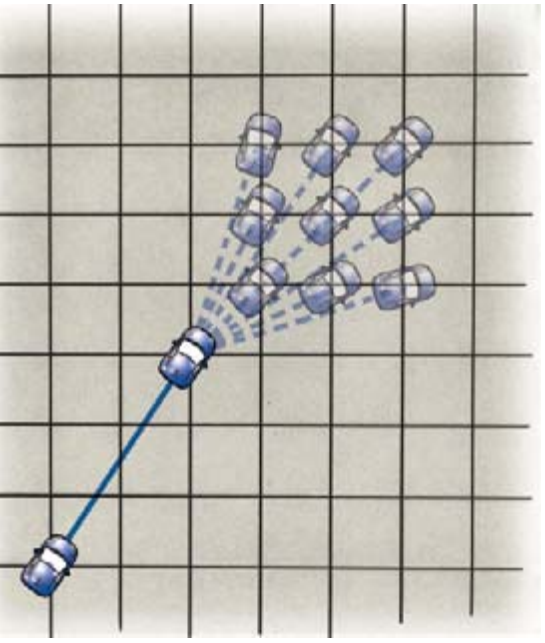
Carreras cuadriculadas

Un conocido y divertido juego, carreras de coches en papel cuadriculado, puede ayudarnos a aprender algo de física

Juan M. R. Parrondo

Hace algunos años, se hizo popular un juego sencillo e ingenioso: carreras de coches en papel cuadriculado. El único material necesario es una hoja de papel cuadriculado y un bolígrafo (o varios de colores, para mayor refinamiento). Puede participar tantos jugadores como deseen.

Se dibuja en el papel el plano de un circuito de carreras. El trazado puede ajustarse a la cuadrícula o no. La pista puede estrecharse, ensancharse, tener largas rectas, curvas cerradas, etcétera. Se elige una línea de salida desde la que se inicia la trepidante carrera. Cada jugador conduce un coche dibujando a trazos su trayectoria por turnos. Los posibles movimientos de un coche en un turno dependen de cómo se movió en el turno anterior. Si realizó un movimiento de coordenadas (n,m), en el siguiente turno podrá realizar el mismo movimiento (n,m), aumentar o disminuir una de estas coordenadas en una unidad como máximo (y siempre que la casilla de destino no esté ya ocupada por otro vehículo).



1. LAS LINEAS DISCONTINUAS muestran los nueve posibles movimientos en un turno, siendo el trazo continuo el movimiento en el turno anterior.

Podemos ver un ejemplo en la figura 1. En un turno, el coche azul se ha desplazado 2 casillas hacia la derecha y 3 hacia arriba. En el turno siguiente, puede moverse a cualquiera de los nueve puntos señalados en la figura.

Esta regla intenta reflejar que los coches tienen una capacidad limitada para cambiar su velocidad, es decir, para acelerar, frenar o girar. La regla cumple sorprendentemente bien su objetivo. En la figura 2 podemos ver una "carrera" entre dos coches, en la que, después de una recta, hay que realizar un giro cerrado. Los coches parten de la línea de salida con poca velocidad y aceleran en la recta. El coche rojo ha elegido bien su estrategia y logra encarar la curva con la velocidad adecuada. Sin embargo, el coche azul ha acelerado demasiado y acaba saliéndose de la pista. (Cuando esto ocurre, en el siguiente turno se vuelve a la carretera con velocidad uno, es decir, adelantando una casilla en cualquier dirección.)

Los trazos remedan bastante bien la trayectoria real de un vehículo. De hecho, para jugar les recomiendo que se dejen llevar por la intuición, puesto que intentar anticiparse a varias jugadas, contando con precisión las casillas que quedan para una curva, por ejemplo, es bastante complicado (y aburrido).

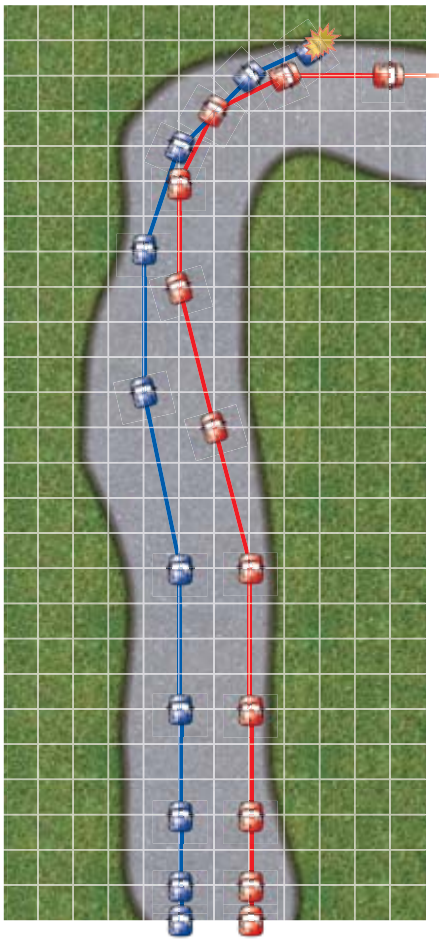
¿Por qué estas reglas tan simples reproducen con algo de realismo el comportamiento de un coche? Si entendemos por "velocidad" del vehículo su desplazamiento por "unidad de tiempo", es decir, por turno, lo que la regla de movimiento limita es la variación de dicha velocidad, que puede como máximo ser de una casilla de las dos direcciones. En otras palabras, la aceleración de nuestro coche no puede ser mayor de una casilla en cada dirección (para ser más precisos, no puede ser mayor que "una casilla por turno al cuadrado"). Y esto es lo que le ocurre a un coche real: su capacidad de acelerar depende de la potencia del motor, la de frenar de la calidad de los frenos y la de girar de la adherencia de los neumáticos. Recordemos que, en la disciplina de la física que se dedica a describir el movimiento de un cuerpo, la cinemática, acelerar, frenar y girar son esencialmente lo mismo: variaciones de la velocidad, es decir, aceleración. Limitando la aceleración a una casilla, conseguimos reproducir de forma aproximada las características del movimiento real del coche.

Las carreras cuadriculadas son, por tanto, una forma divertida de ilustrar muchos conceptos e ideas de la cinemática, como el carácter vectorial de la velocidad y de la aceleración, que a nuestros alumnos les trae de cabeza en el bachillerato e incluso en la universidad. Las trayectorias en la cuadrícula constituyen en realidad una cinemática "discreta", es decir, una cinemática en la que tanto el espacio como el tiempo toman valores enteros: las casillas de la cuadrícula en el caso del espacio y los turnos en el caso del tiempo.

Es instructivo comprobar cómo cambia la cinemática cuando se "discretiza". Sea el movimiento uniformemente acelerado, en el que la aceleración es constante. En la cuadrícula, un movimiento con una aceleración constante de una casilla consistiría en un movimiento en línea recta en el que el coche avanza en cada turno una casilla más. Si en el primer turno avanza una casilla, en el segundo avanza 2, en el tercero 3, etc. El desplazamiento neto después de t turnos sería entonces:

$$d(t) = 1 + 2 + 3 + \ldots + t = t(t+1)/2$$

Los lectores con conocimientos de cinemática recordarán que el desplazamiento para el movimiento "normal" con una aceleración a es $at^2/2$. Las dos



2. UNA "CARRERA CUADRICULADA". El coche azul ha acelerado demasiado en la recta y se ha salido en la curva, mientras que el rojo la ha tomado de forma adecuada.

fórmulas se parecen, para $a = 1$, si t mucho mayor que 1, es decir, tras un gran número de turnos. En este caso, el espacio recorrido es bastante grande y, por ello, el movimiento "discreto", es decir, sobre la cuadrícula, recuerda al movimiento uniformemente acelerado habitual, en el que tiempo y espacio son continuos.

Otro ejemplo clásico de la cinemática elemental nos lo ofrece el movimiento parabólico de un proyectil en la superficie terrestre. El proyectil tiene una velocidad horizontal constante y una velocidad vertical sometida a una aceleración provocada por la gravedad y, por tanto, dirigida hacia abajo. Es sencillo obtener el análogo discreto (*ejemplos de la figura 3*). El desplazamiento horizontal es siempre el mismo, mientras que el vertical disminuye paulatinamente una casilla por turno. No es difícil demostrar que los puntos de cualquier trayectoria de este tipo se encuentran en una parábola real. Los puntos de la trayectoria roja se encuentran en la parábola:

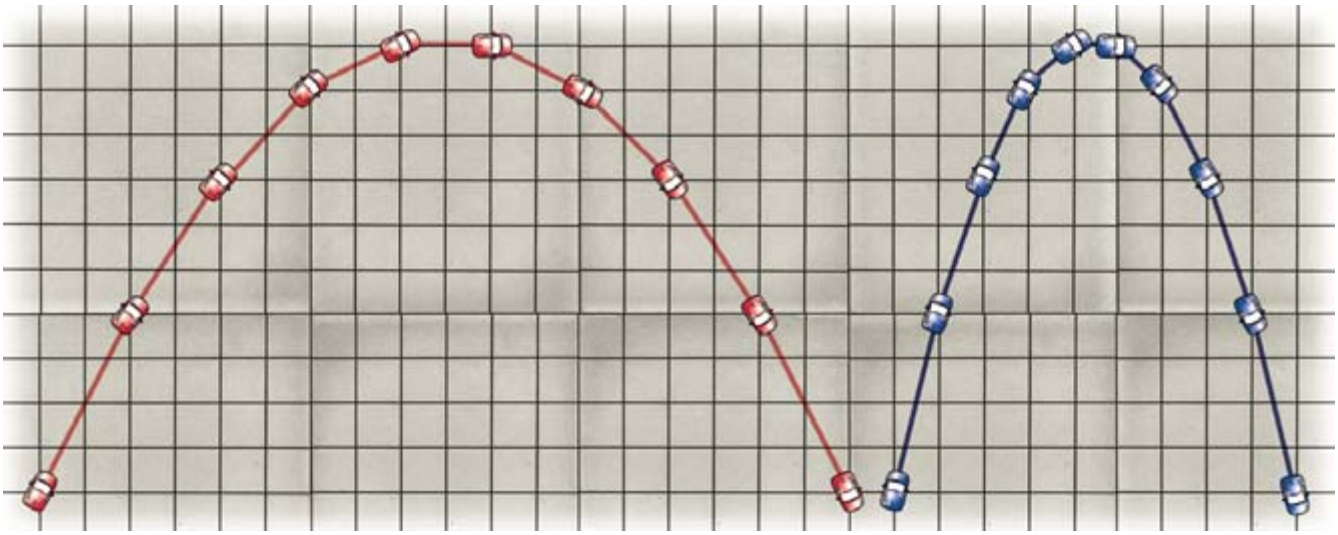
$$y = x(18 - x)/8 = 2{,}25x - x^2/8$$

que es, como pueden comprobar los lectores familiarizados con la cinemática, la trayectoria de un cuerpo que se mueve con velocidad horizontal de dos casillas por segundo, aceleración vertical de una casilla por segundo al cuadrado dirigida hacia abajo y velocidad vertical inicial de 5 casillas por turno. La trayectoria azul está también formada por puntos en una parábola: $y = 4{,}5x - x^2/2$, correspondiente a un móvil con velocidad horizontal de 1 casilla por turno, velocidad vertical inicial, 4,5 casillas por segundo, y aceleración vertical de 1 casilla por segundo al cuadrado.

3. Dos parábolas "discretas": la roja corresponde a un móvil con una velocidad horizontal de dos casillas por turno y la azul, de una casilla por turno.

Aunque las velocidades iniciales varían ligeramente, la versión discreta del movimiento de proyectiles reproduce con bastante fidelidad el movimiento real en espacio y tiempo continuos. Lo que se explica por la sencillez de este tipo de movimiento. Más concretamente, ello se debe al hecho de que la aceleración apunta sobre una de las direcciones de la cuadrícula: la vertical. Si quisiéramos, por ejemplo, reproducir el movimiento de los planetas en torno al Sol, los puntos no estarían exactamente sobre una elipse, sino sólo de forma aproximada. Pero bastaría hacer la cuadrícula más pequeña, o lo que es lo mismo, las velocidades y aceleraciones muy grandes en términos de casillas por turno y por turno cuadrado, respectivamente, para obtener trayectorias cada vez más parecidas a una elipse real. Así acontece en las simulaciones por ordenador de cualquier tipo de movimiento, en películas, videojuegos o aplicaciones científicas. En el ordenador, el movimiento se simula en un espacio y tiempo necesariamente discretos. Pese a todo, resulta asombroso el realismo de tales simulaciones.

Las carreras cuadriculadas pueden considerarse, en cierto modo, un videojuego rudimentario, en el que los cálculos a realizar y la representación gráfica son tan simples, que los mismos jugadores pueden realizarlos con un simple bolígrafo. En cualquier caso, les recomiendo que lo practiquen: no sólo es instructivo; se hace también entretenido.



El caso de la moneda cambiada

Las mismas ideas matemáticas sirven para desenmascarar a un tramposo, interpretar imágenes de satélite o analizar el genoma

Juan M. R. Parrondo

Pablo entró en mi despacho jadeando, despeinado y con el nudo de la corbata aflojado. Sin duda, había tenido un mal día: "¡Me han robado más de tres mil euros! Tienes que ayudarme." Le pedí que se tranquilizara y tomara asiento. "Otra vez mi problema con el juego,... pero esta vez me han hecho trampas, me han engañado de verdad, y te necesito para desenmascarar al estafador. Me propuso jugar con una moneda: si salía cara, ganaba yo 100 euros, y si salía cruz, yo le daba 100 euros a él. Un juego estúpido, pero ya sabes que soy incapaz de decir que no. Probamos la moneda y me convencí de que no estaba trucada. De hecho, en las primeras tiradas gané algo de dinero, pero después cambió mi suerte. El caso es que estoy convencido de que el tipo reemplazó secretamente la moneda en algún momento durante las tres horas que estuvimos jugando. El casino nos cedió una sala y grabó la partida en vídeo, pero el responsable se niega a darme la cinta completa. Dice que únicamente puede dejarme revisar algunos fragmentos. De modo que necesito saber cuándo se hizo el cambio. Y aquí entras tú. Tengo la lista de los resultados de las 200 tiradas en que ha consistido la partida. En algún momento empezaron a aparecer más cruces que caras, pero no soy capaz de averiguar cuándo. Es endiabladamente difícil".

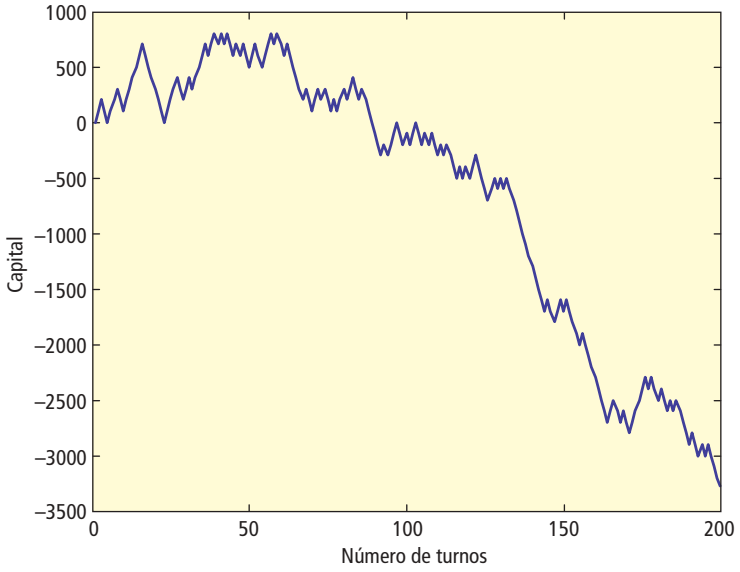
"Déjame esos datos. Veré qué se puede hacer." Pablo alargó una hoja arrugada y llena de números y letras. En ella estaban anotados los resultados de las doscientas tiradas y lo ganado en cada turno. Introdujimos los datos en una hoja de cálculo y comenzamos a analizarlos. La evolución del capital de mi amigo se puede ver en la figura 1. Es difícil determinar en qué momento cambió su suerte. Hay varios períodos de pérdidas sistemáticas, del turno 125 al 140, pero también un poco antes del turno 100 y también alrededor del turno 30. Para solucionar nuestro problema necesitamos un método más riguroso y cuantitativo que la simple inspección visual de la figura.

En esta sección hemos abordado problemas similares (*Fósiles y lotería*, abril de 2005), en los que se dispone de una serie de números o datos aleatorios, pero se desconocen algunos parámetros del procedimiento con el que dichos datos se han obtenido. En este caso, disponemos de los resultados de las tiradas e ignoramos en qué momento se ha cambiado la moneda y cuál es el sesgo de la nueva moneda (aunque esto a mi amigo no le importaba mucho). Uno de los métodos más utilizados para resolver este tipo de problemas es el de *máxima verosimilitud*. Consiste en calcular la probabilidad de que hayan salido los datos, en función de los parámetros desconocidos, y después buscar los parámetros que hacen que esta probabilidad sea máxima. De modo que apliqué la técnica a los datos de mi amigo Pablo.

Supongamos que la moneda inicial es justa y que ha sido reemplazada inmediatamente después del turno N por otra en donde la probabilidad de salir cara es p. Cualquier secuencia de resultados de los N primeros turnos es igualmente probable y su probabilidad es $(1/2)^N$. No ocurre así con los turnos siguientes, en donde la probabilidad de una secuencia de resultados depende del número n de caras que hayan aparecido. La probabilidad de una secuencia dada de resultados es el producto de 200 factores:

$$p_n = \left(\frac{1}{2}\right)^N p^n (1-p)^{200-N-n}$$

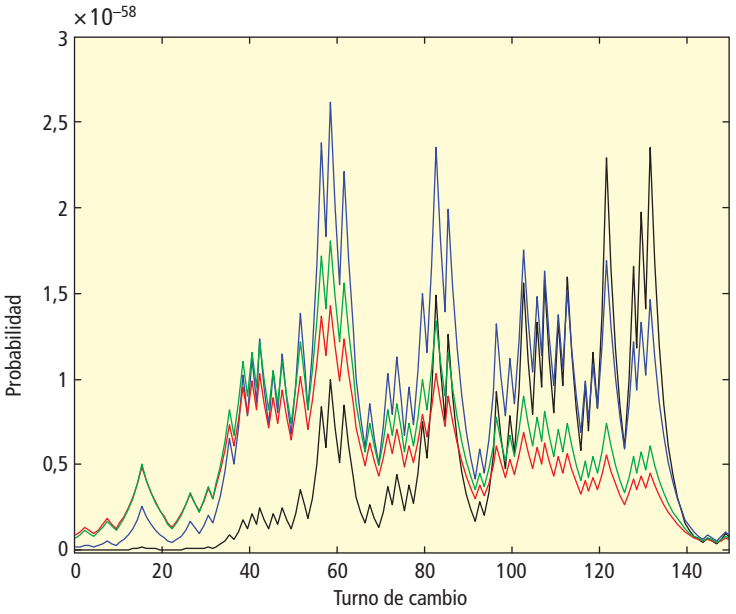
N factores valen 1/2, y corresponden a los resultados de los N primeros turnos jugados con la moneda justa; n factores valen p, que es la probabilidad de que salga cara con la moneda sesgada, y que corresponden a las n caras que han salido con dicha moneda; finalmente, $200-N-n$ es el número de cruces que han aparecido con la moneda sesgada y la probabilidad de que al lanzar dicha moneda salga una cruz es $1-p$. Observen que n depende de nuestro parámetro N puesto que, recordemos, n es el número de caras aparecidas después del cambio de moneda, que se produce en el turno N. Por lo tanto, p_n es una función de los resultados de las tiradas y, lo que es más importante, de los dos parámetros desconocidos: el turno de cambio, N, y el



1. Evolución del capital de mi amigo en función del número de turnos.

sesgo de la moneda trucada, p. En la figura 2 pueden ver el valor de p_n obtenido para los resultados de la partida que jugó Pablo, en función del turno de cambio N y para distintos valores de p.

El procedimiento de máxima verosimilitud nos dice que tenemos que escoger los valores de N y p que hacen máxima esta probabilidad. A la vista de la curva, deducimos entonces que los valores más probables de los dos parámetros son N = 59 y p = 0,35. Aunque hay otros valores con probabilidades cercanas (por ejemplo N = 83 y p = 0,35 o N = 132 y p = 0,3), me arriesgué y le dije a Pablo: "Busca alrededor del turno 59". Reconozco que en la conclusión pesó también un argumento más psicológico que matemático: Pablo iba ganando bastante dinero en el turno 50 y es probable que el rival se pusiera nervioso y quisiera cambiar la moneda cuanto antes.



2. Probabilidad de que se dé la secuencia de datos que realmente ocurrió en la partida, en función del turno en el que se cambió la moneda y para distintos valores del sesgo de la moneda trucada: p = 0,3 (*curva negra*), 0,35 (*curva azul*), 0,39 (*curva verde*) y 0,4 (*curva roja*).

Lo cierto es que Pablo revisó las cintas de vídeo y, en efecto, pudo ver cómo el estafador había cambiado la moneda en el turno 67. A la vista de las pruebas, el tramposo confesó y devolvió el dinero a mi amigo. La moneda trucada tenía en realidad un sesgo p = 0,39. Como pueden observar, la estimación de máxima verosimilitud de N y p no es perfecta: el valor real, N = 67 y p = = 0,39, tiene una probabilidad alta, aunque no máxima. Aun así, el método es bastante bueno, incluso con un reducido número de datos.

Existe otra forma más completa de obtener una estimación de los parámetros desconocidos: la estimación bayesiana, que ya examinamos en otra ocasión (*Fósiles y lotería*, abril de 2005). En la estimación bayesiana se utiliza una cierta distribución de probabilidad *a priori* para los parámetros desconocidos. Por ejemplo, podríamos considerar que N está uniformemente distribuido entre 10 y 150 (el tramposo no se atrevería a cambiar la moneda antes del turno 10, para que Pablo no desconfiara demasiado pronto y, por otro lado, le serviría de poco cambiarla más tarde del turno 150) y p entre 0,2 y 0,45. Como ven, hay cierta arbitrariedad en la elección de la distribución *a priori*, que depende de la información que poseemos acerca de los parámetros o de suposiciones razonables que realicemos sobre sus posibles valores. El método bayesiano permite calcular una nueva distribución de probabilidad incorporando los datos disponibles (en nuestro caso, el resultado de las 200 tiradas). Desgraciadamente, no tenemos aquí espacio para aplicar el método bayesiano a nuestro ejemplo.

¿Qué tiene que ver la historia de mi amigo ludópata con los satélites y la genómica? El enigma que me planteó Pablo es, en realidad, un conocido problema matemático que aparece en diversos campos: el *problema de la frontera*. En efecto, lo que hemos hecho en el caso del cambio de moneda, ha sido determinar la frontera que divide un tipo de comportamiento (la moneda justa) de otro (la moneda sesgada).

El mismo problema aparece cuando se analizan imágenes de satélite y se quiere obtener a partir de ellas y de forma automatizada una clasificación del terreno fotografiado. Imaginen que intentamos averiguar qué zonas del terreno corresponden a bosques y qué zonas corresponden a cultivos. Los píxeles de la imagen son como los resultados de las tiradas en nuestra historia. Su distribución de probabilidad diferirá si el terreno correspondiente a cada píxel es un bosque o un campo cultivado y la estimación de máxima verosimilitud (o la bayesiana) nos permitirá, igual que nos ha permitido averiguar el turno donde se efectuó la trampa, localizar la frontera entre el bosque y el cultivo.

El problema de la frontera aparece también en al análisis de secuencias de ADN. Hoy sabemos que la secuencia de ADN no es una mera concatenación de genes, es decir, de fragmentos que codifican proteínas. Existe el famoso ADN basura, cuya función es aún enigmática, e incluso las partes que sí codifican proteínas tienen una estructura bastante más compleja de lo que se creía: están formadas por fragmentos de gen (los exones) separados por otros (los intrones) que se traducen en ARN pero acaban desechándose en la fabricación de la proteína, mediante cortes y empalmes (un proceso que se denomina *splicing*). Los intrones y los exones son difíciles de distinguir, aunque las secuencias correspondientes tienen composiciones distintas desde un punto de vista estadístico. Identificar dónde acaba un exón y empieza un intrón, o viceversa, es fundamental para interpretar adecuadamente el genoma, es decir, para conocer la composición de las proteínas del organismo. Y, de nuevo, ésta es una tarea similar a averiguar en qué turno el rival de Pablo cambió la moneda.

Tanto en el análisis de imágenes de satélite como el de secuencias de ADN utilizan modelos más complicados que para un simple cambio de moneda. Se llaman genéricamente *modelos de Markov ocultos*. Todos ellos se basan en la misma idea: un sistema genera secuencias de datos (lanzamientos de moneda, píxeles, bases del ADN), pero la estadística de dichas secuencias depende del estado interno del sistema (moneda, origen de la imagen, tipo de fragmento de ADN), que a su vez cambia aleatoriamente.

El juego de las avalanchas

Un modelo matemático para estudiar terremotos y avalanchas se ha transformado en un juego de estrategia

Juan M. R. Parrondo

Supongo que pocos lectores habrán oído hablar de DisX, el juego de las avalanchas. Es un juego de mesa que inventó hace unos años Wolfgang Kinzel, físico de la Universidad de Würzburg. Curiosamente, sólo se comercializa en Israel y en Corea, pero es fácil construirlo uno mismo: sólo se necesitan un tablero cuadriculado de 6 por 6 casillas (una parte de un tablero de ajedrez) y fichas como las de damas (80, aproximadamente).

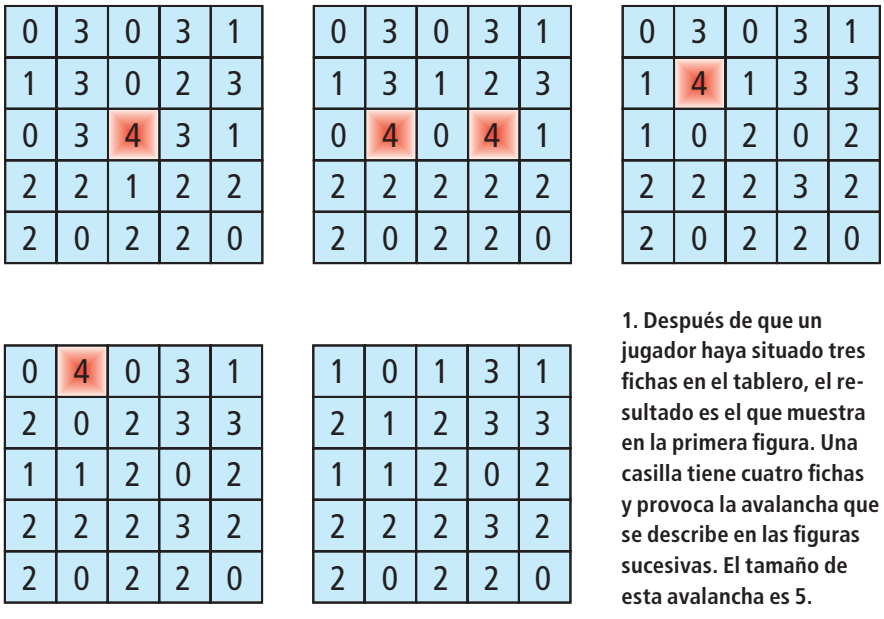
El juego consiste en colocar las fichas en el tablero, apilándolas en las casillas, para intentar provocar avalanchas o reacciones en cadena. La regla fundamental es que no puede haber 4 o más fichas en una casilla. Cuando esto ocurre, la casilla "explota" y sus cuatro fichas se reparten entre las casillas vecinas. Si la casilla que explota está en un borde o en una esquina, se reparten sólo 3 o 2 fichas respectivamente, y se dejan las otras fuera del tablero. Una explosión puede inducir otras, como se muestra en la figura 1. Un jugador ha situado una ficha en la casilla central, que pasa a tener cuatro fichas. La explosión que se produce en esa casilla central genera una serie de explosiones, a la que denominaremos *avalancha*. El tamaño de la avalancha es el número de explosiones que se han producido a raíz de la primera

Cada jugador coloca por turno tres fichas en el tablero. Su objetivo es lograr la mayor avalancha posible, puesto que recibe como puntuación el tamaño de la avalancha producida por sus tres fichas. Se puede comenzar con el tablero vacío o con algunas fichas ya colocadas. En el sitio web de Kinzel (kinzel disx en Google), la partida comienza con las casillas llenas de una y dos fichas de forma alternada. Es una buena elección, que permite provocar avalanchas desde el primer turno. Obsérvese que, si bien cada jugador en su turno coloca tres fichas en el tablero, el número total de éstas no aumenta indefinidamente, puesto que se retiran fichas siempre que hay explosiones en los bordes.

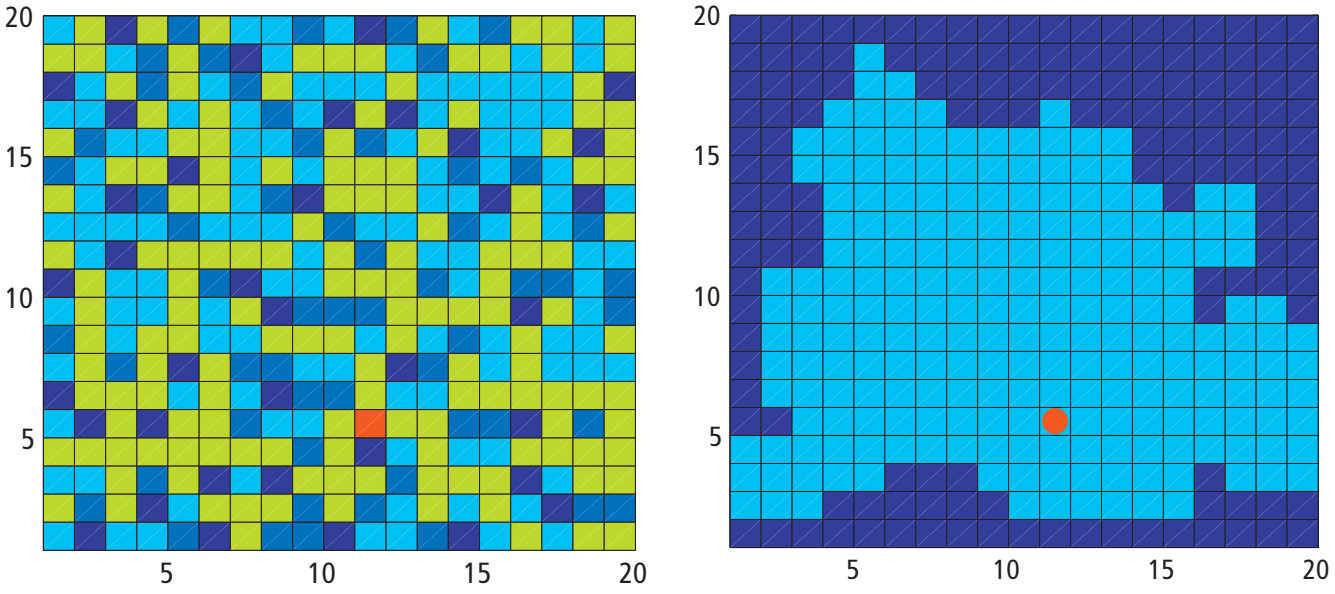
Puede jugarse a DisX por Internet, en la página web de Kinzel, contra un programa de ordenador diseñado por él mismo. Después de una cuantas partidas, nos percataremos de algunas características básicas del juego. La más obvia, la importancia de los sitios con tres fichas, que actúan como "canales" para la avalancha. Lo difícil es utilizarlos de forma óptima y, sobre todo, no dejar muchos de ellos que puedan ser aprovechados después por el adversario.

Sin embargo, la dinámica de las avalanchas es más compleja que una mera propagación a través de sitios de tres fichas. En ocasiones pueden llegar a explotar sitios con una ficha o incluso vacíos, si a lo largo de la avalancha lo hacen sus vecinos. En la figura 2 pueden ver un tablero de 20 × 20, en el que una sola ficha (en rojo) produce una avalancha que ocupa casi todo el tablero. Como pueden ver en las figuras, muchos de los sitios que han explotado estaban vacíos inicialmente (casillas azul oscuro en la figura de la izquierda). Es cierto que las casillas con tres fichas (verdes en la figura de la izquierda) actúan de propagadores de la avalancha, pero resulta difícil predecir el espacio cubierto por ésta (en azul claro en la figura de la derecha), a la vista de la disposición inicial de las fichas.

Para dibujar la figura 2, reproduje en el ordenador el juego de Kinzel, aunque las fichas se fueron colocando al azar hasta que apareció una avalancha de tamaño considerable. Esta versión de DisX, en la que no hay jugadores, sino que se añaden fichas al azar en el tablero, constituye el modelo matemático que inspiró a Kinzel. Lo crearon en 1988 Per Bak, Chao Tang y Kurt Wiesenfeld, del Laboratorio Nacional de Brookhaven. El artículo en el que presentaron su modelo tenía un título un poco críptico: *Criticidad autoorganizada* (*Sel-organized criticality*). A pesar de ello y de la simplicidad del modelo, se trata de uno de los trabajos de los años ochenta más citados en el campo de la física. La criticidad autoorganizada (SOC) puso de manifiesto un comportamiento colectivo que podemos observar en multitud de fenóme-



1. Después de que un jugador haya situado tres fichas en el tablero, el resultado es el que muestra en la primera figura. Una casilla tiene cuatro fichas y provoca la avalancha que se describe en las figuras sucesivas. El tamaño de esta avalancha es 5.



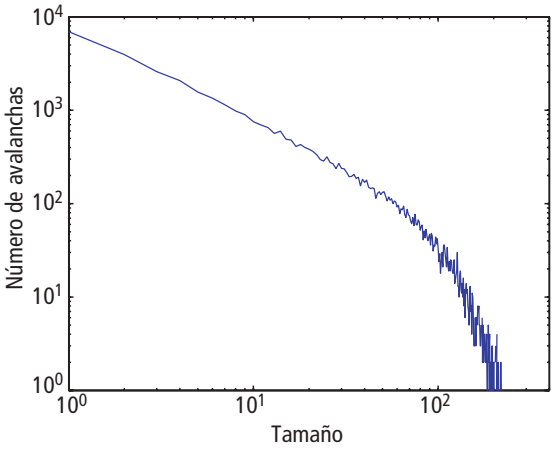
2. En la figura de la izquierda podemos ver el estado del tablero antes de una gran avalancha. Las casillas azul oscuro no tienen ninguna ficha, las azules tienen una ficha, las azul claro 2 fichas y las verdes, 3. Una nueva ficha se añade en la casilla roja, que pasa a tener 4 fichas y provoca una avalancha en la que están involucradas más de la mitad de las casillas. En la figura de la derecha se pueden ver, en color claro, las casillas involucradas en la avalancha.

nos: terremotos, tormentas, pilas de arena e incluso en la actividad neuronal.

Para explicar este comportamiento tan generalizado, he simulado el modelo de Bak-Tang-Wiesenfeld (que, recordemos, no es más que una partida de DisX en un tablero mayor y donde las fichas se colocan al azar) y he registrado el tamaño de las avalanchas que se producen. La distribución de tamaños se puede ver en la figura 3. La forma de la distribución se torna más evidente si se representan los datos en ejes logarítmicos, como se ha hecho en la figura.

Como cabe observar, la distribución sigue una línea bastante recta para tamaños inferiores a 100 casillas y después se curva hacia abajo. Hay un gran número de avalanchas pequeñas, pero se producen de vez en cuando algunas que abarcan casi todo el tablero. Si hiciéramos éste más grande, veríamos que la distribución dibuja una línea recta en un rango mayor de tamaños.

Una recta en ejes logarítmicos significa que las dos magnitudes: tamaño T y número $N(T)$ de avalanchas con dicho tamaño, están relacionadas mediante una *ley de potencias*, $N(T) = AT^{-b}$, de la que hemos hablado en otras ocasiones (*La misteriosa ley del primer dígito*, diciembre, 2002; *Números y palabras,* febrero, 2003). Las leyes de potencias aparecen en muchos fenómenos, desde la frecuencia de uso de palabras hasta la distribución de tamaños de ciudades o intensidades de terremotos.



3. Distribución de tamaños de avalanchas en un tablero de 20 × 20.

¿Puede la *SOC* explicar estas leyes de potencias? La respuesta es: "no todas, pero probablemente algunas de ellas". En concreto, las que exhiben los rasgos fundamentales del modelo de Bak: la existencia de una cierta energía o tensión (las fichas) distribuida en el espacio (el tablero) y que, al superar cierto umbral en una zona (4 fichas en una casilla), hace que esta zona relaje y transmita la tensión a las zonas colindantes. Zonas que pueden, a su vez, superar el umbral y relajar difundiendo la tensión a lo largo de una amplia región del espacio en forma de avalancha.

Es probable que los terremotos se generen mediante un mecanismo similar. El movimiento de las placas tectónicas genera tensión a lo largo de una falla y, si esta tensión supera determinado umbral en algún lugar, el terreno se reajusta, transmitiendo la tensión a las zonas vecinas. Los grandes terremotos son, según esta imagen, poderosas avalanchas, similares a la que se representa en la figura 2. Además, el modelo explica correctamente la distribución de intensidades de los terremotos reales, que obedece a una ley de potencias llamada de Gutenberg- Richter, con un exponente b cercano a 1, similar al que se obtiene en el modelo de Bak.

Es probable que otros ejemplos de leyes de potencias, como la distribución de los tamaños de tormentas, las avalanchas en pilas de arenas o las fracturas en materiales, tengan su origen en la SOC, mientras que la ley de Zipf para la frecuencia de palabras, o leyes de potencias en redes sociales o redes genómicas, necesiten otro tipo de respuesta.

Sorpresas termodinámicas

Si ponemos en contacto dos cuerpos a distintas temperaturas, el más caliente se enfría y el más frío se calienta hasta que las temperaturas de ambos se igualan. Sin embargo, con algo de ingenio se puede desafiar esta afirmación sin contravenir las leyes de la termodinámica.

Juan M. R. Parrondo

Les propongo el siguiente problema. Disponemos de un litro de agua a 90 grados centígrados y de un litro de vino a 10 grados. Supondremos que los dos líquidos tienen el mismo calor específico, es decir, que el incremento o decremento de temperatura en ambos es el mismo cuando se aporta o se extrae la misma cantidad de calor. Si ponemos en contacto el agua y el vino sus temperaturas se igualarán y, como el calor que cede el agua al vino disminuye la temperatura de aquélla en la misma cantidad en que aumenta la de éste, la temperatura final será la media de las temperaturas iniciales, es decir, (90 + 10)/2 = 50. Este estado, en el que las temperaturas se igualan y ya no hay flujo de calor entre los dos cuerpos, se denomina equilibrio termodinámico.

El problema que les propongo es el siguiente: ¿serían capaces de manipular los dos líquidos de modo que el agua termine con una temperatura inferior a la temperatura final del vino? Fijemos con precisión las reglas del juego: pueden poner en contacto una parte de un líquido con una parte del otro, pero no mezclar el agua y el vino; supondremos también que todos los contactos se realizan sin pérdidas de calor, es decir, los líquidos sólo intercambian calor entre sí y no con el ambiente o con los recipientes que los contienen. Esto se podría conseguir mediante un termo con dos receptáculos separados por una pared delgada, que absorba, ceda o absorba y ceda poco calor y sea buena conductora térmica.

Existen muchas soluciones al problema. Tomamos, por ejemplo, medio litro de agua a 90º y lo ponemos en contacto con el litro de vino a 10º. La temperatura final de estos dos líquidos será:

$$\frac{90\times 0{,}5 + 10\times 1}{1{,}5} = \frac{110}{3} \approx 36{,}67^\circ$$

Ahora tomamos el otro medio litro de agua que sigue a 90º y lo ponemos en contacto con el litro de vino a 36,67º. La temperatura final del conjunto es:

$$\frac{90\times 0{,}5 + (110/3)\times 1}{1{,}5} = \frac{490}{9} \approx 54{,}44^\circ$$

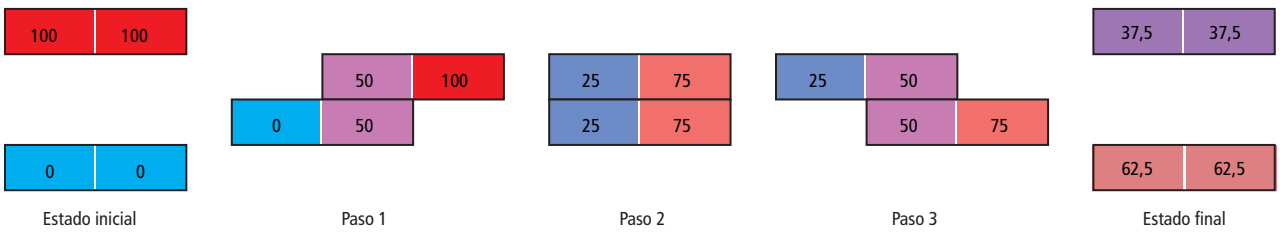
Finalmente, mezclamos el agua: medio litro a 36,67º y otro medio litro a 54,44º. La temperatura de la mezcla será el valor medio de estas dos cantidades:

$$\frac{1}{2}\left(\frac{110}{3} + \frac{490}{9}\right) = \frac{410}{9} \approx 45{,}56^\circ$$

mientras que la temperatura final del litro de vino es 54,44º. Por lo tanto, el agua ha alcanzado finalmente una temperatura inferior a la del vino, a pesar de que inicialmente se encontraba más caliente. La pareja de líquidos se ha "saltado" el estado de equilibrio termodinámico (ambos a 50º). Recordemos que, según la segunda ley de la termodinámica, es necesario aportar energía para que un sistema se aleje del equilibrio (para ser más precisos, hay que aportar energía libre). Nuestro ejemplo no viola la segunda ley, a pesar de que en ningún momento hemos aportado energía al conjunto de los dos líquidos desde el exterior. La razón es que el conjunto nunca pasa por el estado de equilibrio: sí lo hace el vino, que en algún momento del proceso ha estado a 50º, pero no el agua, que en ningún momento está, en su totalidad, a 50º.

El truco de poner en contacto sólo parte de los líquidos puede dar más de sí. Para verlo, imaginemos ahora dos sólidos del mismo material y masa y con forma de barra, uno a 100º y otro a 0º. Para poder realizar los contactos parciales, supondremos que la difusión del calor a lo largo de cada barra es mucho más lenta que el flujo de calor entre una barra y otra cuando están en contacto. Veamos lo que ocurre si procedemos tal y como se muestra en la figura 1. En primer lugar, ponemos en contacto dos mitades de cada una de las barras. Estas dos mitades se equilibrarán a una temperatura de 50º, mientras que las mitades que no están en contacto continuarán con las temperaturas iniciales. En el segundo paso, ponemos en contacto toda la barra. La mitad a 50º de la barra superior entra en contacto con la mitad a 0º de la inferior y ambas se equilibran a 25º. Por otro lado, la mitad a 50º de la barra inferior entra en contacto con la mitad a 100º de la superior y ambas se equilibran a 75º. Desplazamos de nuevo las barras en el tercer paso poniendo en contacto la mitad superior a 75º y la inferior a 25º, que se equilibran a 50º. Finalmente, separamos las dos barras y esperamos a

1. Deslizando un cuerpo sobre otro, conseguimos enfriar el cuerpo inicialmente caliente por debajo de los 50º y calentar el frío por encima de dicha temperatura.

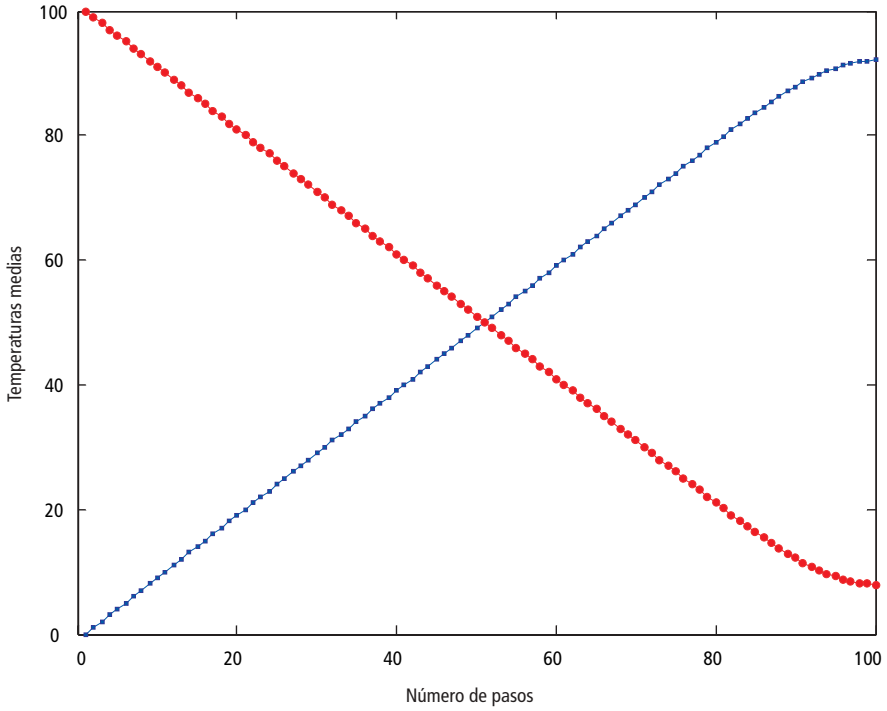


2. Generalización del proceso de la figura 1, en donde las barras se dividen en un número arbitrario de fragmentos.

que la temperatura se haga uniforme en cada una de ellas. La superior, una de cuyas mitades está a 25º y la otra a 50º, alcanzará la temperatura media de 37,5º, mientras que la barra inferior se equilibrará a 62,5 . (E lector habrá observado que tanto en el primer ejemplo de los líquidos como en éste, la suma de las temperaturas finales es igual a la suma de las iniciales: este hecho es consecuencia del principio de conservación de la energía.)

Se podría realizar un proceso similar con el agua y el vino de nuestro primer ejemplo, consiguiendo temperaturas finales de 60º para el vino y 40º para el agua. Este método se muestra, pues, más eficaz para intercambiar temperaturas (recordemos que nuestra primera solución al problema conducía a temperaturas finales de 54,44º y 45,56º, respectivamente). ¿Existe un procedimiento que permita intercambiar completamente las temperaturas, es decir, que el agua acabe a 10º y el vino a 90º o, en el caso de las barras, la superior a 0º y la inferior a 100º? La respuesta es afirmativa. Con más precisión, podemos acercarnos cuanto queramos al intercambio perfecto de temperaturas.

3. Evolución de las temperaturas medias de las dos barras cuando éstas se dividen en 50 fragmentos.



Para lograrlo, basta generalizar el proceso de la figura 1, dividiendo las barras en muchos fragmentos en lugar de sólo dos. Por ejemplo, si dividimos las barras en 50 partes y deslizamos una barra a lo largo de la otra, las temperaturas finales resultan ser: 92,0411º para la barra inferior (inicialmente a 0º) y 7,9589º para la superior (inicialmente a 100º). El proceso con 50 fragmentos consta de 100 pasos, puesto que hay que deslizar la barra inferior, tal y como se muestra en la figura 2.

En la figura 3 se ofrece la evolución de las temperaturas medias de cada barra a lo largo del proceso. Recordemos que la temperatura en cada barra no es uniforme, salvo inicialmente y cuando, después de separar las barras, dejamos que se equilibren. Por lo tanto, esta temperatura media sólo es significativa al final y al comienzo del proceso (aunque es proporcional a la energía interna de cada barra).

La suma de las temperaturas medias es constante. Por la sencilla razón de que la energía total del sistema se conserva. En el paso 50, cuando las dos barras están alineadas, las dos temperaturas medias son iguales, ya que las dos barras están equilibradas fragmento a fragmento. Finalmente, si aumentamos el número de fragmentos, nos acercamos más al completo intercambio de temperaturas. Por ejemplo, para 10.000 fragmentos, las temperaturas finales son 99.4358º y 0.5642º. El comportamiento de las temperaturas medias es siempre similar: son casi líneas rectas durante todo el proceso salvo en los últimos pasos. A pesar de ello, no he sido capaz de encontrar una expresión matemática para estas temperaturas (las gráficas están hechas con un sencillo programa informático). Ni siquiera he podido demostrar que, cuando el número de fragmentos es infinito, las temperaturas se intercambian, aunque intuyo que existe un argumento de simetría relativamente sencillo ¿Puede el lector dar con él?

El lector con conocimientos de física se habrá percatado también de un hecho sumamente interesante: cuando las temperaturas se intercambian de forma perfecta, la entropía del sistema no crece. A pesar de que las dos barras tienen al principio temperaturas muy diferentes y han intercambiado calor durante todo el proceso, ¡este es obviamente reversible si las temperaturas finales son iguales a las iniciales!

Los procesos descritos en las figuras 1 y 2 se pueden interpretar como el deslizamiento de las barras en sentidos contrarios. Una idea similar se utiliza en el diseño de algunos intercambiadores de calor, dispositivos con numerosas aplicaciones, sobre todo en sistemas de refrigeración, desde aparatos de aire acondicionado hasta el radiador de un coche o los refrigeradores de motores y calderas. Los intercambiadores de calor más eficientes son precisamente aquellos en los que un fluido frío y otro caliente circulan por tuberías adyacentes y en sentido contrario, algo que recuerda bastante al deslizamiento de las barras que se muestra en la figura 2. Incluso los delfines utilizan este método para minimizar las pérdidas de calor cuando están en aguas muy frías: en el sistema circulatorio, algunas arterias por las que circula sangre caliente proveniente del corazón se hallan rodeadas de capilares con sangre fría procedente de la parte más exterior del animal. El intercambio de calor se produce en esa red capilar, de modo que la sangre que regresa al corazón está caliente y la que llega a la parte exterior del animal está a temperatura baja, disipando poco calor al exterior.

Estimaciones

En muchas ocasiones nos tenemos que contentar con respuestas aproximadas. Pero un buen argumento estimativo puede enseñarnos más que la solución exacta de un problema

Juan M. R. Parrondo

¿Cuántos afinadores de piano hay en Chicago? Aunque les cueste creerlo, éste es un problema "clásico" en el gremio de los físicos. No porque haya teorías físicas acerca de la afinación del piano (que las hay), sino porque esta misma pregunta la lanzaba con asiduidad Enrico Fermi, uno de los físicos más relevantes de la primera mitad del siglo XX, a sus alumnos de la Universidad de Chicago. Lo que Fermi pretendía con ello era enseñar una forma de razonamiento extremadamente útil en física y otras ciencias: las estimaciones. Es decir, el intento de abordar un problema no con el objetivo de dar una respuesta exacta, sino sólo aproximada. En este intento no necesitamos conocer todos los datos que conducen a la solución exacta (algo que muy raramente ocurre en ciencias). Podremos hacer hipótesis "razonables", aunque tengamos escasas pruebas de su validez.

Veamos cómo resolvía Fermi su clásico problema. En Chicago hay unos 5 millones de habitantes. En cada hogar hay, de media, dos personas; de modo que tenemos unos 2 millones de hogares. Supongamos que hay un piano en uno de cada 20 hogares (este dato, desgraciadamente, ya ha dejado de ser cierto incluso en orden de magnitud). Hay entonces en Chicago unos 100.000 pianos, que suelen afinarse una vez al año. Por otro lado, podemos suponer que un afinador es capaz de afinar un piano en unas dos horas, incluyendo el transporte. Afinará, en una jornada de ocho horas, unos 4 pianos. Si trabaja 5 días a la semana durante 50 semanas al año, será capaz de afinar, $4 \times 5 \times 50 = 1000$ pianos al año. Por tanto, la afinación de los 100.000 pianos de Chicago requerirá unos 100 afinadores.

Fíjense en la cantidad de hipótesis que hemos realizado a lo largo del argumento, algunas no tan razonables para nuestra época, aunque sí lo eran en los EE.UU. de los años cuarenta. Incluso hemos cometido errores matemáticos de bulto, como afirmar que la mitad de 5 millones es aproximadamente 2 millones. En las estimaciones, todas estas hipótesis y cálculos aproximados son aceptables, puesto que lo que nos interesa es sólo el *orden de magnitud* de la respuesta final. Hemos obtenido 100 afinadores de piano, lo que significa que el número real puede estar, digamos, entre 50 y 150. A pesar de la inexactitud del resultado, la información que aportan las estimaciones es muy valiosa, sobre todo si no hay otra forma de obtener una respuesta más precisa o si para ello se requiere un gran esfuerzo económico o de cálculo. Así lo pensaba Fermi, quien consideraba las estimaciones esenciales en la formación de los físicos. En su honor, los problemas que se abordan mediante métodos estimativos se conocen en muchos foros como *problemas de Fermi.*

Una de las primeras estimaciones célebres fue la de la población de la ciudad de Londres, realizada por el comerciante John Graunt en 1662 a partir de datos de nacimientos y muertes y reseñada recientemente por el historiador I. B. Cohen en su libro *El triunfo de los números*. En la época de Graunt no había un censo de la población total, pero ya se contaba con registros fiables del número de muertes al año y del número de bautizos (muy ligeramente inferior al de nacimientos). En el Londres de finales del siglo XVII, se registraba una media de 12.000 nacimientos al año. Graunt supuso que las mujeres londinenses en edad fértil tenían un hijo cada dos años, lo cual conduce a la cifra de 24.000 mujeres en edad fértil. Suponiendo que el período de fertilidad abarca desde los 16 hasta los 40 años, mientras que la edad de las mujeres casadas varía entre los 16 y los 76 años, concluyó que había:

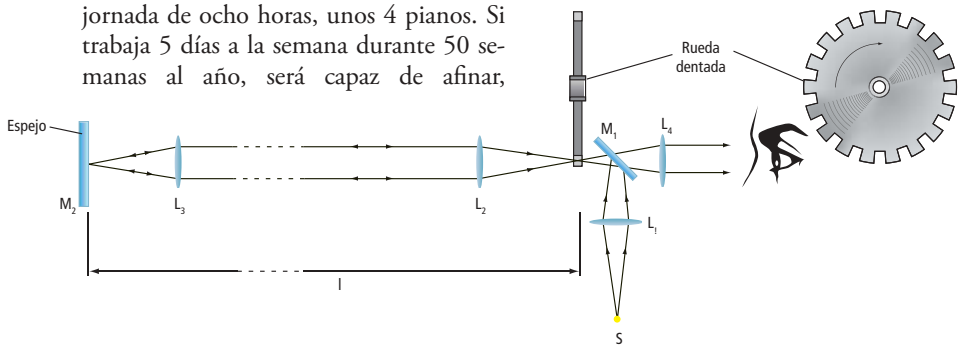
$$24.000 \times \frac{76-16}{40-16} = 24.000 \times \frac{60}{24} \approx 48.000$$

mujeres casadas. Aquí vemos ya una aproximación considerable, puesto que 60 no es el doble de 24. Probablemente Graunt realizó esta corrección para tener en cuenta la pirámide poblacional, más estrecha en las edades avanzadas. Finalmente, Graunt observó que cada unidad familiar londinense tenía una media de 8 miembros: la madre, el padre, tres hijos y tres sirvientes o huéspedes. Por lo tanto, la población total londinense tenía que ascender a unos $48.000 \times 8 = 384.000$ personas.

Graunt confirmó este resultado por otros dos métodos. En primer lugar, observó, a partir de datos de parroquias pequeñas, que se producían de media tres muertes al año por cada 11 familias. Como el número de fallecimientos en todo Londres ascendía a 13.000 anuales, el número total de familias debía ser:

$$13.000 \times \frac{11}{3} \approx 48.000$$

en pleno acuerdo con la estimación basada en los nacimientos. La segunda comprobación se basaba también en el número de muertes. A partir de la densidad de



1. Esquema del experimento de Fizeau para determinar la velocidad de la luz.

población, estimó que el número de familias que habitaban dentro de las murallas de la ciudad era de 12.000. Por otra parte, sabía que, de todas las muertes que se producían, dos tercios ocurrían extramuros. Por tanto, el número total de familias debía de ser tres veces el número de familias intramuros, es decir 12.000 × × 3 = 48.000.

Los historiadores de la ciencia consideran a Graunt el "padre" de la estadística matemática, por estas estimaciones publicadas en 1662 en un pequeño libro titulado *Observaciones naturales y políticas mencionadas en índice adjunto y basadas en las listas de mortalidad*. Sus contemporáneos también saludaron la obra de Graunt como una importante contribución científica y, a pesar de que poseía escasa formación académica, fue elegido miembro de la Regia Sociedad tras la publicación de su libro.

En el caso de Graunt, la estimación era el único modo de abordar el problema de la población de Londres. En otras ocasiones, las estimaciones pueden servir para diseñar un experimento que luego nos dará una medida precisa de cierta cantidad o para analizar su viabilidad. Son fascinantes los métodos de medida desarrollados para obtener cantidades muy pequeñas o muy grandes, especialmente en sus inicios. ¿Cómo podemos conocer el tamaño de un átomo o el número de moléculas en un gramo de agua? ¿De qué manera una distancia de 10^{-10} metros se traduce en algo mensurable macroscópicamente? Los experimentadores tienen que encontrar la forma de que distancias, tiempos, masas o fuerzas muy pequeñas den lugar a efectos macroscópicos. Veamos un ejemplo.

La luz viaja a unos 300.000 kilómetros por segundo. Ya Rømer, a finales del siglo XVII, obtuvo una estimación bastante precisa de esa cifra observando la diferencia en los períodos de los satélites de Júpiter cuando la Tierra se acerca o se aleja del planeta. Si sabemos que la velocidad es del orden de los 300.000 kilómetros por segundo y queremos realizar una medida directa de dicha velocidad, ¿qué estrategia deberíamos seguir?

Para una medición directa, es decir, una medición del tiempo empleado por la luz en recorrer una cierta distancia, o bien utilizamos distancias muy grandes, como hizo Rømer, y tiempos apreciables, o bien utilizamos distancias cortas y formas muy precisas de medir el tiempo. Fizeau llegó a un compromiso entre estas dos estrategias y realizó el primer experimento de medición directa de la velocidad de la luz en 1849. La idea consiste en enviar pulsos de luz a cierta distancia en donde se coloca un espejo, detectar el pulso que retorna y medir el tiempo empleado en el viaje. Mediante lentes focalizadoras, Fizeau logró enviar el rayo a algo menos de 10 kilómetros de distancia y detectar con precisión el rayo reflejado. El viaje total de la luz sería entonces de 20 kilómetros. Para medir la velocidad de la luz, necesitaríamos entonces medir tiempos del orden de

$$\frac{20}{300.000} \sim 10^{-4}\,\text{segundos}.$$

¿Era posible medir con tanta precisión un intervalo de tiempo, con los medios de la primera mitad del siglo XIX, puramente mecánicos? Fizeau diseñó un método ingenioso para lograrlo. Construyó un disco dentado con 720 muescas en su circunferencia (*véase la figura*). Podía hacer girar el disco a velocidades bien controladas, hasta unas 15 revoluciones por segundo. Por lo tanto, a esta velocidad máxima, el tiempo entre una hendidura y la siguiente es de

$$\frac{1}{720 \times 15} \approx 10^{-4}\,\text{segundos}.$$

¡Ya tenemos la precisión temporal deseada! El esquema completo del experimento se puede ver en la figura 1. El espejo semirreflectante M_1 envía la luz al espejo lejano, pero permite que la luz reflejada lo traspase. El experimentador observa el rayo que retorna y empieza a girar la rueda dentada a velocidades cada vez mayores hasta que la luz desaparece. Ello se debe a que el rayo en su viaje de ida se encuentra un hueco de la rueda y al volver un diente. Fizeau observó que la luz se desvanecía cuando la rueda giraba a 12,1 revoluciones por segundo. Por lo tanto, el tiempo empleado por la luz en su viaje de ida y vuelta era

$$\frac{1}{12{,}1\ \text{rev/s} \times 2 \times 720} \approx 5{,}74 \times 10^{4}\,\text{segundos}$$

(recordemos que hay 720 dientes, pero 2×720 dientes más muescas). En su experimento, la distancia entre la rueda y el espejo lejano era de 8,630 kilómetros. Por lo tanto, la velocidad de la luz resulta:

$$c = \frac{2 \times 8{,}630\ \text{km}}{5{,}74 \times 10^{4}\ \text{s}} \approx 300.738\ \text{km/s}.$$

2. Enrico Fermi (1901-1954).

Combinando las dos fórmulas anteriores, es fácil ver que la velocidad c aparece como el producto de tres factores medibles: la velocidad de rotación de la rueda, el número de dientes y muescas y dos veces la distancia entre la rueda y el espejo lejano. Gracias a este producto, podemos llegar al orden de magnitud de la velocidad de la luz, que es de 10^8 metros por segundo, mediante factores relativamente "asequibles": distancias del orden de los 10 kilómetros (10^4 metros), un millar de dientes y huecos, y velocidades angulares en torno a las 10 revoluciones por minuto. Se puede decir que Fizeau apuró cada uno de los factores al máximo de las posibilidades experimentales de su época. La medida final no está exenta de error, pues hay cierta subjetividad en la determinación de la velocidad a la que la luz que retorna es bloqueada por el diente de la rueda. Sin embargo, el experimento de Fizeau fue bastante preciso (el valor aceptado de c hoy en día es 299.792,458 m/s).

Lo importante es que Fizeau no pudo embarcarse en la ardua tarea de construir su dispositivo experimental sin antes hacer una estimación del número de dientes, velocidad de la rueda y longitud del recorrido de la luz, necesarios para determinar c. Probablemente un argumento puramente estimativo, utilizando el valor previo de c debido a Rømer, fue su guía para diseñar el experimento.

Cifras y letras

Un programa informático resuelve, con bastante ingenio, los problemas numéricos y verbales del conocido concurso televisivo

Juan M. R. Parrondo

La mayoría de los lectores conocerán el concurso *Cifras y Letras*. En él se alternan dos pruebas, una numérica y otra de construcción de palabras. En la primera de ellas se dan seis números, que pueden ser 1, 2, 3,... 9, 10, 25, 50, 75 o 100, y un número "objetivo", que se encuentra entre 101 y 999. Con los seis primeros números hay que obtener, mediante las cuatro operaciones elementales (suma, resta, multiplicación y división), el número objetivo o acercarse a él lo máximo posible.

A pesar de la simplicidad matemática del problema, plantea algunas cuestiones interesantes. Cuando el número objetivo es grande, siempre me hago la siguiente pregunta: "¿Cuál es el número máximo que puede obtenerse a partir de los seis números iniciales?" La respuesta no es difícil. Entre dos números cualesquiera, la multiplicación es la operación que da mayores resultados, excepto si uno de los números es 1, en cuyo caso es mejor sumar. La estrategia idónea para conseguir el mayor número posible es entonces multiplicar los seis números iniciales entre sí, si ninguno de ellos es 1. Si ocurre esto último, ¿cómo deberíamos sumar el 1 para maximizar el resultado? Como $(a + 1) \times b = a \times b + b$, habrá que conseguir que b sea lo más alto posible. Por tanto, la estrategia óptima consiste en sumar el 1 al menor de los números iniciales y multiplicar el resultado por los cuatro números restantes. Por ejemplo, el mayor número que se puede conseguir con 1, 3, 2, 10, 5 y 8 es $10 \times 8 \times 5 \times 3 \times (2 + 1) = 3600$.

Pero volvamos al problema original del concurso: conseguir el número objetivo a partir de los seis números dados. Pedro Reina, profesor de matemáticas, ha diseñado un programa capaz de resolver el problema en menos de un segundo, es decir, capaz de encontrar la combinación de operaciones que más se acerca al número objetivo. Podemos estar seguros de que se trata de la mejor combinación, porque el algoritmo que utiliza el programa es una búsqueda exhaustiva a través de todas las posibles combinaciones. El método para codificar y explorar estas combinaciones, de notable ingenio, ilustra una de las estrategias típicas de la inteligencia artificial: la búsqueda en árboles de decisiones.

Reina ha conseguido organizar todas las posibles combinaciones en un árbol que se crea a partir de una serie de decisiones, tal y como se muestra en la figura 1. Cada decisión consiste en elegir dos números y una operación a realizar entre ambos. No hay ninguna ambigüedad con respecto al orden en el que realizar la operación, puesto que la suma y la multiplicación son conmutativas y la resta y división sólo pueden dar un número entero y positivo con un orden dado. De hecho, para muchos pares de números, la división no puede realizarse de modo que el resultado sea un número entero, en cuyo caso dicha operación queda excluida.

En el primer ejemplo de la figura 1, hemos tomado el 10 y el 7 y los hemos multiplicado. Eliminamos de la lista los dos números seleccionados, el 10 y el 7, y añadimos el resultado de la operación, en este caso 70. Por tanto, tras cada paso, tenemos una serie de números más reducida. Después de cinco pasos el resultado es un único número. El algoritmo de Reina explora todas estas posibles decisiones hasta encontrar el número objetivo en cualquiera de los pasos dados. Por ejemplo, en el segundo caso mostrado en la figura 1, el número objetivo, 288, se ha encontrado en el cuarto paso, por lo que no es necesario continuar. El algoritmo almacena el número que más se ha acercado al objetivo a lo largo de la exploración; por tanto, si en la búsqueda no se encuentra finalmente el número exacto, el programa da como solución la aproximación más cercana.

¿Es posible esta exploración exhaustiva en un tiempo razonable? Para responder a esta pregunta, calculemos de entrada el tamaño del árbol de combinaciones. En el primer paso podemos elegir $(6 \times 5)/2 = 15$ parejas de números y aplicarles 4 operaciones. Suponiendo que la división es siempre aplicable, el número de opciones en el primer paso es $(4 \times 6 \times 5)/2 = 60$. En general, si disponemos de n números, las posibles opciones son:

$$4 \times n(n-1)/2 = 2n(n-1)$$

Por tanto, el número total de opciones en los cinco pasos es:

$$2^5 \times 6 \times (5 \times 4 \times 3 \times 2)^2 = 2.764.800$$

es decir, algo menos de 3 millones de posibilidades. Si descartamos la división, el resultado sería:

$$(3/2)^5 \times 6 \times (5 \times 4 \times 3 \times 2)^2 = 656.100$$

ligeramente superior al medio millón de posibilidades. La cifra final debe, pues, hallarse entre estas dos cantidades. En cualquier caso, se trata de un número asequible para que un ordenador moderno realice la exploración exhaustiva en segundos.

Reina ha incluido también en su algoritmo algunos trucos que lo aceleran,

como empezar siempre operando con los números mayores o guardar combinaciones de cuatro números y evitar explorar ramas idénticas del árbol. En la página web pedroreina.net, hay una versión del programa en línea, que tarda pocos segundos, aun cuando se introduzcan números que no tienen solución exacta, situación en la que la exploración debe recorrer todo el árbol de posibilidades.

El programa de Reina no imita el modo como razonamos los humanos. Ante el problema de las cifras, creo que la gente utiliza en paralelo dos estrategias. En una de ellas tratamos de aproximarnos primero de forma "gruesa" al objetivo, combinando los números iniciales grandes y utilizando sobre todo multiplicaciones, para luego afinar la aproximación con los números pequeños mediante sumas y restas. Por ejemplo, en el caso de la figura 1, multiplicaríamos 75 × 4 = = 300 y luego trataríamos de conseguir un 22 mediante combinaciones con el resto de los números. Esta estrategia no conduce, aquí, a la solución exacta.

La segunda estrategia, más elaborada, consiste en descomponer en factores el número objetivo y tratar de construir dichos factores. Por ejemplo, 288 es 72 × 4 y el 72 es relativamente fácil de obtener a partir de los números iniciales. Observen que cada estrategia sitúa al principio o al final la multiplicación, que es una operación clave en todo el juego, ya que posee un mayor abanico de resultados.

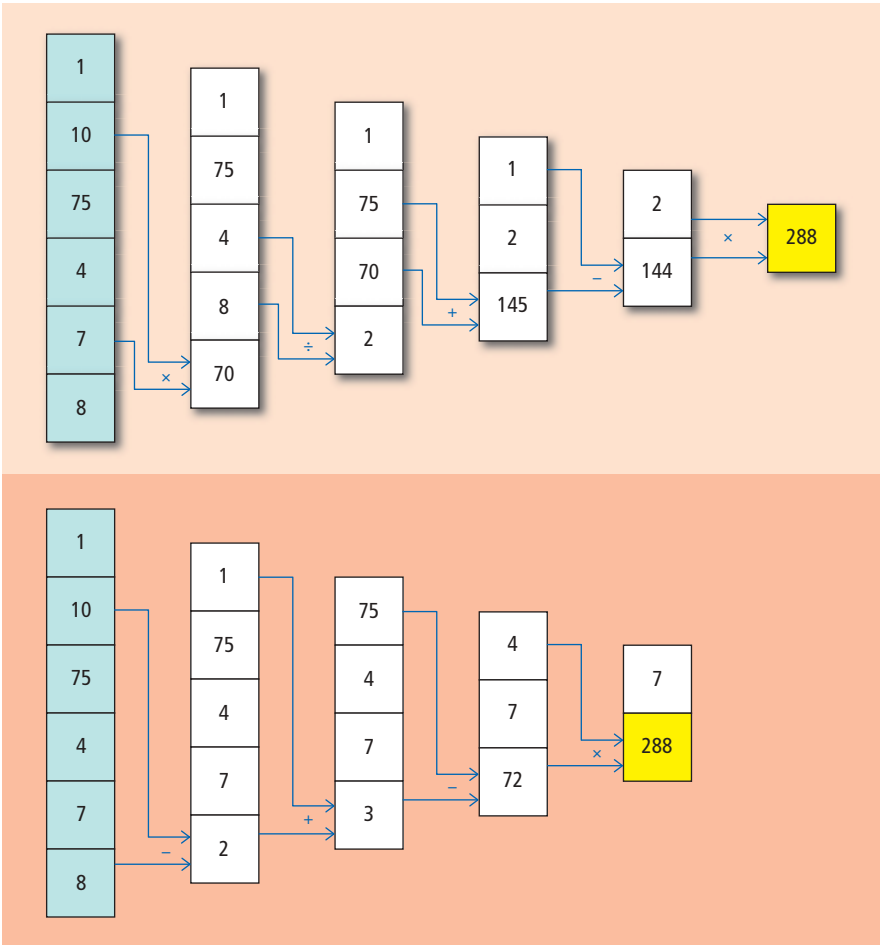
Programar estas estrategias es muy complicado, porque no podemos formularlas con precisión y en ellas interviene en buena medida la intuición. Es como si recorriéramos el árbol de posibilidades de forma un tanto errática, comenzando a veces por el final y otras por el principio, descartando ramas por pura sospecha y adentrándonos con profundidad en otras que nos parecen prometedoras. Esta exploración tan "humana" supone un reto para la inteligencia artificial. En el problema de las "Cifras", la exploración exhaustiva es posible, pero en otros más complejos la máquina sigue estando por detrás de la mente. Lograr algoritmos de exploración que remeden la intuición del ser humano sigue siendo uno de los principales retos de la inteligencia artificial.

Reina ha diseñado también un algoritmo capaz de resolver la prueba de las "Letras". El problema es muy simple: se dan nueve letras y los concursantes han de generar con ellas la palabra más larga posible.

La solución que nos ofrece Reina es muy sencilla, gracias a una idea sumamente ingeniosa que el propio autor atribuye al profesor de inteligencia artificial Antonio Salmerón. Al escribir el programa, toma primero todas las palabras del diccionario de la Real Academia y ordena alfabéticamente las letras de cada palabra. Por ejemplo, "nutritivo" se convierte en "iinorttuv". Agrupa entonces estos anagramas según su longitud, los ordena alfabéticamente y los guarda en nueve ficheros: uno para los anagramas de nueve letras, otro para los de ocho, y así sucesivamente. Esta ordenación de las palabras del diccionario es la clave para encontrar la solución al problema.

Cuando nos dan las nueve letras con las que construir una palabra, el algoritmo las ordena alfabéticamente y busca en el fichero de anagramas de nueve letras. La búsqueda es muy rápida porque los anagramas están ordenados alfabéticamente. Si no encuentra la solución, elimina una de las letras y busca en el fichero de anagramas de ocho letras. Tiene que probar eliminando todas las letras, lo que da lugar a 9 posibilidades. Si no encuentra la solución, procede a eliminar dos letras y así sucesivamente. Como vemos, la idea de guardar las palabras del diccionario en forma de anagramas es clave para no tener que probar permutaciones de letras. Con ello se consigue un algoritmo bastante rápido y exhaustivo, como se puede comprobar en la página web de Reina.

En este caso, la solución dada, a pesar de su ingenio, no tiene nada que ver con el razonamiento humano. La forma en que las personas resolvemos este problema es aún más compleja que en el caso de las "Cifras" y utiliza una capacidad genuinamente humana: el uso del lenguaje. Y son precisamente los procesos cognitivos relacionados con el lenguaje natural los que suponen un mayor desafío en el campo de la inteligencia artificial.



Dos posibles soluciones de un problema de "Cifras", en el esquema de árbol de Pedro Reina. En el segundo caso, se llega al número deseado sólo en cuatro pasos.

Encuestas electorales

Entre lo que declaran los ciudadanos en las encuestas electorales y la estimación de voto que hacen las empresas demoscópicas hay grandes diferencias. ¿Cómo se obtiene la estimación del voto a partir de las respuestas obtenidas en las encuestas?

Juan M. R. Parrondo

Un par de semanas antes de las últimas elecciones generales del 9 de marzo, casi todos los periódicos publicaron en primera página los resultados del sondeo preelectoral realizado por el Centro de Investigaciones Sociológicas (CIS) que auguraba un "empate técnico" entre los dos grandes partidos políticos, PSOE y PP, asignando al primero un 40,2 % de los votos válidos y al segundo un 38,7 %. Sin embargo, al leer el estudio con más detalle, se podía observar que a la pregunta "¿si las elecciones fueran mañana, a qué partido votaría usted?" un 31,0 % de los encuestados respondía PSOE, mientras que un 21,1 % se decantaba por el PP. Añadiendo a estos datos las simpatías mostradas por los aún indecisos, el PSOE alcanzaba un 37,1 % y el PP un 24,5 %. ¿Cómo es posible que una diferencia de casi 10 puntos en la intención de voto, o de más de 12 puntos en la intención de voto más la simpatía, se tradujeran finalmente en una de 1,3, es decir, en un "empate técnico" (bastante cercano a los resultados finales)?

En el mismo estudio del CIS, antes de dar paso al anexo que contiene la estimación de voto, se advierte: "*Dado que los datos de los indicadores "intención de voto" e "intención de voto + simpatía" son datos directos de opinión y no suponen ni proporcionan por sí mismos ninguna proyección de hipotéticos resultados electorales, en este anexo se recogen los resultados de aplicar un modelo de estimación a los datos directos de opinión proporcionados por la encuesta. Obviamente, la aplicación a los mismos datos de otros modelos podría dar lugar a estimaciones diferentes.*"

¿Por qué la intención de voto no es una buena estimación del voto que se producirá en las elecciones? ¿Y cuál el modelo utilizado por el CIS para estimar el voto a partir de los datos de la encuesta? Aunque dicho modelo no se publica en la página *web* del CIS, el director de su departamento de investigación sí ha respondido amablemente a mis consultas y me ha proporcionado información detallada acerca del procedimiento de estimación.

El problema principal con el que se enfrentan los técnicos del CIS es la gran, y constante, discrepancia entre el recuerdo de voto en las pasadas elecciones de 2004 y el voto real que se produjo en aquellos comicios. En la Tabla I se pueden ver las respuestas a la pregunta "¿a quién votó en las últimas elecciones generales (2004)?" junto con los porcentajes de voto que realmente tuvieron lugar. Es difícil explicar estas discrepancias, que además se producen en multitud de encuestas, no sólo en las de intención de voto. ¿Se deben a una elección incorrecta de la muestra o a que algunos encuestados no dicen la verdad?

Aunque es poco probable que sea así, lo más sencillo desde un punto de vista teórico es suponer que la discrepancia se debe a un error muestral. Esta suposición permite corregir los datos de cualquier pregunta realizada en la encuesta. Para ello hace falta cruzar las respuestas de todas las preguntas con el recuerdo de voto, tablas que están disponibles en la página *web* del CIS. Si p_j es el porcentaje de individuos que recuerdan haber votado a j y c_{ij} es, de entre estos últimos, el porcentaje de individuos que se decantan por una opción i en una pregunta, el porcentaje total de sujetos que se decantan por i en la encuesta será:

$$c_i = c_{i1}\frac{p_1}{100} + c_{i2}\frac{p_2}{100} + \ldots$$

Sin embargo, si el porcentaje real de votos en 2004 por el partido j fue $\tilde{p}_j$, entonces, una buena estimación del porcentaje real de individuos que se decantan por la opción i en la pregunta será:

$$\tilde{c}_i = c_{i1}\frac{\tilde{p}_1}{100} + c_{i2}\frac{\tilde{p}_2}{100} + \ldots$$

en lugar de c_i. Esta corrección equivale a reajustar la muestra para que refleje correctamente el recuerdo de voto. Para aplicar la fórmula anterior a los datos de la encuesta electoral, hay que tener en cuenta que ciertas opciones en la pregunta de recuerdo de voto no tienen reflejo en el voto real en 2004 y, al revés, no se pregunta si el voto del encuestado fue nulo. Por tanto, los porcentajes $\tilde{p}_j$ no son exactamente los porcentajes de voto de 2004, sino que hay que incluir aquellos encuestados que no tenían edad para votar, no recuerdan el voto o no contestan a la pregunta del recuerdo de voto en 2004 y hay que eliminar los votos nulos de los porcentajes de voto real. La siguiente fórmula refleja estas modificaciones:

$$\tilde{p}_j = \frac{100 - P(n.e.) - P(n.r.) - P(n.c.)}{100 - P(v.n.)}\, P_j(v.r.)$$

en donde P denota los porcentajes de voto nulo real en 2004 (*v.n.*), no tenía edad para votar (*n.e.*), no recuerda (*n.r.*), no contesta a la pregunta del recuerdo de voto (*n.c.*), y $P_j(v.r.)$ es el porcentaje de votos que consiguió el partido j en 2004. En la tercera columna de la Tabla I se muestran los porcentajes $\tilde{p}_j$ calculados con la fórmula anterior.

Vayamos ahora a las preguntas de la encuesta que se refieren a la intención de voto. Por razones de espacio, limitaré todo el análisis a los tres partidos más votados (en el 2004 y el 2008): PSOE, PP e IU, pero los lectores pueden completar el mismo estudio con el resto de partidos utilizando los datos cruzados que están a su disposición en la página *web* del CIS.

En la Tabla II se puede ver lo que se denomina intención de voto más simpatía por un partido. Al encuestado se le pregunta en primer lugar "si las eleccio-

	Recuerdo de voto 2004	Voto real 2004	Porcentajes corregidos
PSOE	37.7	32.22	26.69
PP	20.0	28.53	23.63
IU/ICV	3.2	3.75	3.11
CiU	1.8	2.44	2.02
ERC	0.9	1.91	1.58
PNV	0.8	1.23	1.02
BNG	0.4	0.61	0.51
CC	0.3	0.69	0.57
EA	0.1	0.23	0.19
CHA	0.1	0.27	0.23
Na-Bai	0.1	0.18	0.15
Otro partido	1.0	1.74	1.44
En blanco	2.3	1.18	0.98
No votó	13.6	24.34	20.16
No tenía edad	4.8		4.8
No recuerda	4.9		4.9
N.C.	8.1		8.1
Votos nulos		0.76	

Intención + simpatía → / ↓ Recuerdo	PSOE	PP	IU	Abstención
PSOE	76.5	6.2	2.9	11.2
PP	3.7	84.1	0.3	9.8
IU	13.4	1.4	74.1	7.5
CiU	6.1	3.4	1.6	11.7
ERC	7.3	0	5.8	7.5
PNV	3.4	0	0.4	9.5
BNG	15.9	0.9	2.3	13.0
CC	6.8	15.6	4.4	18.7
EA	0	1.9	0	19.4
CHA	32.5	5.6	0	37.1
Na-Bai	3.7	0	0	18.5
Otro partido	6.1	7.8	5.4	17.0
No tenía edad	35	24.9	5.7	27.6
En blanco	13.2	10.3	1.8	65.9
No votó	22.7	17.1	3.4	50.5
No recuerda	20.2	16	2.2	53.4
N.C.	9.2	7.3	1.5	77.7

nes fueran mañana, ¿a qué partido votaría?". El resultado se denomina intención de voto. A los que a esta pregunta responden que votarían en blanco, no votarían o no saben o no contestan, se les pregunta de nuevo "¿por cuál de los siguientes partidos o coaliciones siente usted más simpatía o cuál considera más cercano a sus propias ideas?". La intención más simpatía por un determinado partido se determina sumando las respuestas a favor de dicho partido en las dos preguntas. Los resultados se pueden ver en la Tabla II, cruzados con el recuerdo de voto. Los porcentajes están siempre calculados sobre el total de respuestas de recuerdo de voto por una determinada opción. Por eso las columnas no suman cien. Las filas sí deberían sumar 100, si incluyéramos todos los partidos sobre los que se pregunta. La columna "abstención" incluye aquellos que, en las dos preguntas, no se han decantado por ningún partido, es decir, lo que en la segunda han insistido en que no tienen preferencia por ninguno o han vuelto a marcar la casilla "no sabe" o "no contesta".

© iStockphoto/NOMBRE DEL AUTOR DE LA FOTO

Disponemos ahora de todos los datos necesarios para estimar el voto en las elecciones. Para calcular el voto a IU, por ejemplo, tomamos cada entrada de la tercera columna de la Tabla II y la multiplicamos por el correspondiente porcentaje corregido de la Tabla I. Sumamos todos estos productos y obtenemos el porcentaje de voto final. Lo que hacemos así es aplicar la segunda fórmula de este artículo. Si se procede de esta forma con los tres principales partidos y con la abstención, se obtienen los siguientes resultados: PSOE 30,39 %, PP 27,98 %, IU 4,60 %, abstención 27,62 %. Finalmente, si queremos dar las estimaciones en porcentajes sobre voto válido, tendremos que eliminar la abstención, es decir, multiplicar los valores anteriores por 100/(100 – 27,62). Con ello se obtienen los siguientes resultados: PSOE 41,99 %, PP 38,66 %, IU 6,35 %.

El algoritmo completo que utiliza el CIS tiene algunas correcciones adicionales que utilizan otros cruces y que aquí no hemos incluido. Por eso la estimación final del CIS (PSOE: 40,2 %, PP: 38,7 %, IU: 5,8 %) difiere ligeramente de la calculada aquí. Sin embargo, como pueden comprobar, la corrección por recuerdo de voto es la más importante de todas las que realiza el modelo del CIS.

Sigue siendo un misterio la discrepancia entre el recuerdo de voto declarado y el voto real en 2004. Probablemente se deba la suma de varios factores: algunos desajustes en la elección de la muestra o sesgos en la disponibilidad de las personas a contestar, respuestas no sinceras, o respuestas sinceras pero equivocadas. Estas últimas no deberían en realidad corregirse, puesto que el encuestado puede ser sincero en su respuesta sobre intención de voto, aunque sea algo olvidadizo. Las dos primeras razones de discrepancia sí se corrigen adecuadamente con el método que utiliza el CIS, puesto que si un individuo miente acerca de lo que votó en 2004, lo más probable es que también lo haga sobre su intención de voto. En cualquier caso, la corrección realizada resulta pertinente, como lo demuestran los resultados de las elecciones de este año (PSOE: 43,64 %, PP: 40,11 %, IU: 3,8 %). Las discrepancias entre la estimación y el resultado final pueden deberse perfectamente, además de a errores estadísticos, al desarrollo de la campaña electoral, ya que la encuesta se realizó entre el 21 de enero y el 4 de febrero de 2008.

El problema de los tres dioses

Resumen de la sección, resumen de la sección, resumen de la sección, resumen de la sección, resumen de la sección

Juan M. R. Parrondo

En 1996, George Boolos, profesor de lógica del MIT, publicó en *The Harvard Review of Philosophy* el problema de los tres dioses, una variante de los famosos acertijos de escuderos mentirosos y caballeros sinceros que pueblan los excelentes libros de Raymond Smullyan. Boolos lo presentó, con algo de exageración, como "el problema lógico más difícil de la historia". Veremos primero una variante del mismo y, más tarde, la versión original.

Tenemos ante nosotros tres dioses. Sabemos que uno de ellos es el dios de la Verdad, quien, obviamente, siempre dice la verdad. Otro de ellos es el dios de la Mentira, que siempre miente. Y el tercer dios es el dios de la Confusión, que responde al azar a cualquier pregunta que se le formule.

El problema consiste precisamente en averiguar quién es quién formulando sólo tres preguntas cuya respuesta sea necesariamente sí o no. No valen preguntas del tipo "¿que respondería tu compañero de la izquierda si le preguntara si es cierto que dos más dos son cuatro?", ya que si el dios interrogado es el de la Verdad o el de la Mentira y el compañero de la izquierda fuera el de la Confusión, entonces el interrogado tendría que responder "no sé" o morderse la lengua. Basta con que exista la posibilidad de contestar algo distinto a "sí" o "no" para que la pregunta no sea válida.

Vamos a acotar aún más las preguntas: sólo se puede preguntar por el nombre de los dioses o combinaciones lógicas de los mismos. Se pueden formular las tres preguntas al mismo dios o a dioses diferentes y una pregunta puede depender de las respuestas obtenidas anteriormente. Pero sólo disponemos de tres preguntas y, tras ellas, deberemos saber con certeza el nombre de cada uno de los dioses.

La teoría de la información nos dice que las preguntas óptimas son aquellas en las que las posibles respuestas tienen la misma probabilidad. En nuestro caso, tenemos que hacer preguntas cuya respuesta sea "sí" con probabilidad 1/2 y "no" con la misma probabilidad. De este modo, la pregunta nos aporta un bit de información. Sin embargo, en nuestro problema este criterio no basta (por algo es el problema "más difícil de la historia"), aunque ayuda bastante. No basta por culpa del endiablado dios de la Confusión, cuyas respuestas obviamente no aportan información alguna.

Llamemos A, B y C a los tres dioses que están ante nosotros. Hay seis combinaciones posibles de "personalidades", todas ellas igual de probables: VMC, VCM, MCV, MVC, CMV, CVM. Supongamos que formulamos a A una pregunta óptima, en el sentido de la teoría de la información. Esta pregunta no podrá eliminar en ningún caso las dos últimas posibilidades, CMV, CVM, en las que A es el dios de la Confusión y puede, por tanto, responder cualquier cosa. Si la pregunta es óptima, eliminará sólo dos opciones (la mitad de las cuatro primeras) y nos quedaremos con cuatro posibilidades.

Estas cuatro posibilidades suponen una incertidumbre de dos bits y nos quedan dos preguntas más. Por tanto, las dos últimas preguntas han de ser también óptimas, pero esta vez de verdad. La segunda tiene que eliminar dos posibilidades de las cuatro restantes y, la última, conducirnos a la única instancia posible. Pero hemos visto que el dios de la Confusión no nos proporciona información alguna. Por lo tanto, es condición necesaria para resolver el problema que, tras la primera pregunta, sepamos con certeza que alguno de los dioses no es el de la Confusión.

Esta es la clave: la primera pregunta debe conducirnos a la conclusión de que alguno de los dioses no es el de la Confusión. ¿Cómo podemos conseguirlo? Vimos que, al preguntar a A, las posibilidades CMV y CVM son compatibles con cualquier respuesta y que, si hacemos la pregunta óptima, la respuesta sólo podrá eliminar dos de las posibilidades restantes: MVC, MCV, VMC, VCM. Si queremos, tras la primera pregunta, saber con certeza que alguno de los dos dioses no es el de la Confusión, necesariamente esta pregunta debe eliminar el par VMC, MVC (con lo que sabríamos que B no es el dios de la Confusión) o el par VCM, MCV (con lo que sabríamos que C no es el dios de la Confusión). Una vez determinada la información que nos tiene que proporcionar la pregunta, no es difícil formularla: "dios A, ¿es cierta al menos alguna de estas dos afirmaciones: a) B es el dios de la Mentira y C el dios de la Confusión o b) B es el dios de la Confusión y C el de la Verdad?"

Si A es el dios de la Verdad, sólo la afirmación a) puede ser cierta, y la respuesta será "sí" sólo si la instancia VMC es la verdadera y "no" si la disposición de los dioses es VCM. Si A es el dios de la Mentira, dirá "sí" sólo si las dos afirmaciones son falsas, es decir, sólo si la disposición de los dioses es MVC, y dirá "no" si la disposición es MCV. Por lo tanto, la respuesta "sí" nos deja como instancias posibles VMC y MVC (junto con las inevitables CMV, CVM), mientras que la respuesta "no" es compatible sólo con VCM y MCV, junto con CMV y CVM. Si la respuesta es "sí", sabemos con certeza que B no es el dios de la Confusión. Si

la respuesta es "no", es C quien con toda seguridad no es el dios de la Confusión.

Las dos preguntas restantes son muy sencillas. Si la respuesta a la primera ha sido "sí", preguntamos a B: "¿eres el dios de la Confusión?" Si dice que sí, entonces B tiene que ser el dios de la Mentira, y si dice que no, será el dios de la Verdad. Finalmente, le preguntamos a C por la personalidad de alguno de sus compañeros y con ello resolvemos completamente el problema. La dificultad estaba, sin duda, en la primera pregunta. Debido al "ruido" que produce el dios de la Confusión, la primera pregunta no aporta un bit de información. Si fuera así, permitiría eliminar la mitad de las instancias, es decir, tres de las seis posibilidades iniciales. Sólo elimina dos de ellas, como hemos visto, pero, correctamente formulada, permite que las dos preguntas siguientes sí aporten un bit cada una, reduciendo las cuatro posibilidades que quedan a una sola.

El "problema lógico más difícil de la historia" es, en realidad, un poco más difícil. En la formulación original de Boolos, los dioses responden en una lengua que nosotros desconocemos. Las palabras para "sí" y "no" son "da" y "ja", pero no sabemos cuál es la afirmativa y la negativa. Sin entender siquiera lo que nos responden los dioses, ¡este problema sí parece el más difícil de la historia! Sin embargo, en el problema se elimina la restricción que antes impuse a las preguntas. Ahora no tienen por qué ser únicamente preguntas directas acerca de la personalidad de los dioses.

Hay una forma de obtener información incluso sin saber el significado de las respuestas. Si queremos saber la respuesta a cualquier pregunta *q*, basta preguntar: ¿Responderías "ja" a la pregunta *q*? Supongamos que *q* es cierta. Si el dios dice la verdad y "ja" es "sí", la respuesta será "ja". Pero si "ja" es "no", el dios de la Verdad también responderá "ja", diciéndonos "no responderé 'no' a la pregunta *q*". Es decir, el dios de la Verdad responde "ja" si *q* es cierta y "da" si *q* es falsa, independientemente de lo que signifique "ja". Pero aún hay más. El dos de la Mentira también responde "ja" si *q* es cierta, independientemente de lo que signifique "ja". En efecto, si "ja" es "sí", el dios de la Mentira, que respondería "da" a *q*, respondería "sí" a nuestra pregunta, es decir, respondería "ja". Y si "ja" es "no", el dios de la Mentira, que respondería "ja" a *q*, respondería "no" a nuestra pregunta, es decir "ja".

Por lo tanto, si el dios al que preguntamos no es el de la Confusión, conoceremos si *q* es verdad o no con una simple pregunta. Este truco nos permite resolver el problema con cierta facilidad. La primera pregunta al dios A sería: "¿Si te preguntara "es B el dios de la Confusión", responderías "ja"?". Si la respuesta es "ja", entonces, o bien A es el dios de la Confusión (posibilidad que ninguna respuesta puede eliminar) o bien B es el dios de la Confusión. En cualquier caso, sabríamos que C *no es* el dios de la Confusión y le preguntaríamos sobre la personalidad de A y B. Por otro lado, si la respuesta a nuestra primera pregunta es "da", entonces o bien A es el dios de la Confusión o bien B no es el dios de la Confusión. En cualquier caso, B no sería el dios de la Confusión (puesto que si lo es A, B no puede serlo), y podremos preguntar a B sobre la personalidad de los otros dos.

Como pueden comprobar, con las tres preguntas resolvemos el problema, identificando la personalidad de los tres dioses, sin necesidad de conocer el significado de "da" y "ja". Ni siquiera sabemos su significado después de las tres preguntas. De hecho, si el problema consistiera en adivinar las personalidades de los tres dioses *y* el significado de "da" y "ja", ¡no tendría solución! La razón es que en este caso hay 12 posibilidades iniciales, las 6 disposiciones de personalidades de los tres dioses combinadas con las dos posibilidades para los significados de "da" y "ja". Cada pregunta, si es óptima, elimina la mitad de las posibilidades. Por tanto, con tres preguntas podríamos resolver sólo problemas con un máximo de 8 posibilidades iniciales.

Existen aún más sutilezas y variantes del problema. Por ejemplo B. Rabern y L. Rabern han puesto de manifiesto que es necesario aclarar el comportamiento del dios de la Confusión: ¿responde al azar (es decir, lanza una moneda para decidir su respuesta) o adopta al azar la disposición a decir la verdad o la mentira (es decir, lanza una moneda para decidir si dice la verdad o si miente)? En el segundo caso, el problema es trivial puesto que, como hemos visto, la respuesta "ja" a la pregunta "¿Responderías "ja" si te preguntara *q*?" nos dice que *q* es cierta independientemente de si el dios preguntado es el de la Verdad o el de la Mentira. Por lo tanto, si el dios de la Confusión simplemente actúa como el de la Verdad o como el de la Mentira en cada pregunta, también al responder "ja" nos asegura que *q* es cierta.

Los mismos autores exploran preguntas autorreferentes que algún dios no puede contestar. Por ejemplo: "¿vas a contestar "ja" a esta pregunta que te estoy haciendo?". Si "ja" significa "no" y le hacemos esta pregunta al dios de la Verdad, no podrá responder nada, puesto, que tanto si responde "ja" como "da", estará mintiendo. Como los dioses son infalibles, Rabern y Rabern concluyen que la única salida del dios de la Verdad es que le explote la cabeza. Proponen nuevos problemas en los que se pueden utilizar preguntas que manden la cabeza del dios por los aires. Ahora caben tres posibles respuestas: "ja", "da" y "explosión de cabeza". Por tanto, cada pregunta aporta más de un bit de información.

Si permitimos volar la cabeza de los dioses mediante contradicciones de este tipo, el problema original de Boolos se puede resolver con sólo dos preguntas. Pero esta vez tendrán que esperar un mes para conocer la solución.

Piensa un número

Un truco de adivinación de números de Lewis Carroll nos sirve de excusa para reflexionar acerca de la "irrazonable efectividad de las matemáticas en las ciencias naturales"

Juan M. R. Parrondo

A: "Piensa un número."
B: [Piensa el 23].
A: "Multiplícalo por 3. ¿El resultado es par o impar?"
B: [Obtiene 23 x 3 = 69]. "Es impar."
A: "Súmale 31 o 35, lo que prefieras, y divídelo entre 2."
B: [Suma 31 y obtiene (69 + 31)/2 = 50].
A: "Multiplícalo por 3 y dime si el resultado es par o impar."
B: [Obtiene 50 x 3 = 150]. "Es par."
A: "Súmale 80 o 100, lo que prefieras, y divídelo entre 2."
B: [Suma 100 y obtiene 250/2 = 125].
A: "Ahora añádele el número que pensaste al principio seguido de una cifra cualquiera."
B: [Añade al 23 un 7 y obtiene 237 + + 125 = 362].
A: "Divide el resultado entre 7, desechando el resto."
B: [Obtiene 51].
A: "Vuelve a dividir entre 7, desechando el resto, y dime el resultado."
B: "El resultado es 7."
A: "Pues bien, el número que pensaste es 23."

Este pequeño diálogo es una versión, corregida y simplificada, de una nota que escribió Lewis Carroll en febrero de 1896, dos años antes de su muerte. Sorprendentemente, a pesar de lo intrincado de las operaciones y de la arbitrariedad de algunas de ellas, es posible siempre reconstruir el número original a partir de la información que proporciona B: la paridad de dos de los resultados parciales y la cifra final.

Observen que, para que el truco funcione, es necesario que el "ruido" introducido cuando B suma cantidades elegidas por él (31 o 35 para resultados impares, 80 o 100 para los pares) o cuando añade una cifra cualquiera al número original, no haga que se pierda a lo largo del cálculo la información acerca del número pensado inicialmente. Esto probablemente no extrañe a muchos lectores. Las últimas divisiones entre 7 hacen desaparecer el efecto de estas sumas. Veámoslo en un ejemplo similar. Tomen un número cualquiera x; si añadimos una cifra y a la derecha, el número que resulta es $10x + y$; sumando este número al inicial obtenemos $11x + y$. Si ahora dividimos por 11 despreciando el resto, volvemos a obtener x, puesto que y es menor que 11. La adición de la cifra desconocida es un cierto "ruido" que, sin embargo, desaparece en el cómputo final, debido a que hemos amplificado mucho la "señal" x antes de añadir dicho ruido.

El método de adivinación de Carroll se basa, en buena medida, en esta propiedad matemática, pero es menos trivial, ya que utiliza tanbién información acerca de la paridad de los resultados parciales del cálculo. Se puede comprobar que estas paridades dependen sólo de cuán lejos está de ser múltiplo de 4 el número original.

Podemos ver en la tabla siguiente cómo evoluciona el cálculo según sea el número inicial x un múltiplo de 4 más 1, 2, 3 o 4. En la tabla he indicado en amarillo los resultados impares y en rojo los pares, en los pasos en que la paridad se revela a A. Muestro los posibles resultados después de cada paso. Si éstos son más de dos, especifico, separados por comas, el máximo y mínimo de los posibles valores. Por ejemplo, el sexto paso (sumar el número original x con una cifra añadida y, es decir, sumar $10x + y$), aplicado a un número entre $9n + 109$ y $9n + 134$ (cuarta columna $x = 4n + 4$), puede dar, como mínimo, $9n + 109 + 10(4n + 4) = 49n + + 149$ (es el caso en que $y = 0$) y, como máximo, $9n + 134 + 10(4n + 4) + 9 = = 49n + 183$ (es el caso en que $y = 9$).

Como ven, el resultado final, junto con la paridad de los resultados parciales, determina unívocamente el número original. Para el elegido por B, 23, el resultado ha sido: impar, par y 7. Por lo tanto, nos encontramos en la tercera columna ($x = 4n + 3$), n será $7 - 2 = 5$ y el número pensado $4 \times 5 + 3 = 23$.

La versión original del Carroll es diferente y además incorrecta, según ha señalado y corregido el matemático R. F. McCoart. Mi versión es también ligeramente distinta de la propuesta por McCoart. No

x	$4n + 1$	$4n + 2$	$4n + 3$	$4n + 4$
Multiplica por 3	$12n + 3$	$12n + 6$	$12n + 9$	$12n + 12$
Regla par/impar	$6n + 17$ o $6n + 19$	$6n + 43$ o $6n + 53$	$6n + 20$ o $6n + 22$	$6n + 46$ o $6n + 56$
Multiplica por 3	$18n + 51$ o $18n + 57$	$18n + 129$ o $18n + 159$	$18n + 60$ o $18n + 66$	$18n + 138$ o $18n + 168$
Regla par/impar	$9n + 41$, $9n + 46$	$9n + 80$, $9n + 97$	$9n + 70$, $9n + 83$	$9n + 109$, $9n + 134$
Suma $10x + y$	$49n + 51$, $49n + 65$	$49n + 100$, $49n + 126$	$49n + 100$, $49n + 122$	$49n + 149$, $49n + 183$
División por 7	$7n + 7$, $7n + 9$	$7n + 14$, $7n + 18$	$7n + 14$, $7n + 17$	$7n + 21$, $7n + 26$
División por 7	$n + 1$	$n + 2$	$n + 2$	$n + 3$

ha sido fácil encontrar esta versión simplificada. Hay que elegir cuidadosamente los números que se suman para que el resultado final y la paridad de los resultados parciales determinen unívocamente el número pensado.

Consideren ahora este otro "truco" de adivinación. Piense un número x de 50 a 150. Deje caer una piedra desde esa altura, medida en centímetros, al suelo y compute el tiempo que tarda en caer. Deje caer ahora la piedra desde un metro de altura y mida de nuevo el tiempo. Divida el primero de los tiempos entre el segundo y dígame el resultado y. El número pensado será $x = y^2$. La razón estriba en lo siguiente: la altura desde la que se deja caer un objeto es proporcional al cuadrado del tiempo que tarda en caer. Observen que el truco no se basa en un mero cómputo del tiempo empleado, es decir, en haber medido con anterioridad los tiempos de caída desde todas las posibles alturas. De hecho, nuestro método de adivinación funcionaría en la superficie de cualquier planeta.

En este ejemplo, nuestra capacidad adivinatoria no es consecuencia únicamente de las reglas de manipulación de símbolos matemáticos, sino también de cómo caen los cuerpos en la superficie de un planeta, obedeciendo las leyes de Newton. De todas las relaciones posibles entre altura h y tiempo t de caída, la naturaleza ha elegido que h sea proporcional a t^2. ¿Es realmente una elección arbitraria? ¿Podría existir un universo en donde h fuera proporcional a t^3? La aspiración de muchos físicos es precisamente demostrar que las leyes que rigen el comportamiento del universo son las únicas posibles, determinadas por pura coherencia lógica. Una variante de esta idea ha dado lugar a grandes avances científicos. Por ejemplo, las geometrías no euclídeas son lógicamente coherentes pero, cuando se "inventaron", no había prueba alguna de que pudieran existir en la naturaleza espacios no euclídeos de dimensión mayor que 2. Alguien pudo haberse preguntado entonces: ¿por qué la naturaleza eligió la geometría euclídea y cómo sería el mundo si la elección hubiera sido otra? Años más tarde, Einstein demostró que el espacio deja de ser euclídeo en los lugares donde hay un campo gravitatorio.

Si hubiera una necesidad lógica en que la altura h fuera proporcional a t^2, podríamos decir que la "adivinación newtoniana" sería similar a la "adivinación carrolliana", aunque más compleja y con la lógica interna del mundo físico. Pero, ¿por qué el universo tendría que ser lógico o disponer de una estructura matemática? En otras palabras, ¿por qué la matemática es tan eficaz para describir el mundo físico, hasta el extremo de poder determinar unívocamente su comportamiento?

Las posturas dentro del mundo de la física y la matemática ante estas preguntas son múltiples. Algunos piensan que el universo tiene una estructura intrínsecamente matemática. Otros, siguiendo a Arthur S. Eddington, opinan que la estructura matemática proviene de nuestra forma de percibir y pensar el mundo. Eddington afirmaba que toda la física se podía deducir de forma lógica, analizando las estructuras de nuestra percepción y nuestro pensamiento. Para ilustrar su argumento utilizaba el ejemplo siguiente: unos hombres fueron a pescar al mar con una red; después de examinar lo que habían capturado en varios lugares y condiciones, concluyeron que los peces tenían un tamaño mínimo de cinco centímetros. Sin embargo, para llegar a esa conclusión no hace falta salir en barca ni echar las redes, sino sólo medir los agujeros de la red (que tienen, obviamente, 5 centímetros de lado). Del mismo modo, para deducir toda la física, bastaría, según Eddington, analizar la forma en que percibimos y categorizamos el mundo.

Eugene Wigner sostenía una postura más modesta. No pensaba que todos los aspectos del mundo físico pudieran ser descritos de forma matemática, pero sí veía en las matemáticas una herramienta fundamental para revelar la verdadera naturaleza del universo. Escribió en 1960 un famoso artículo, *La irrazonable eficacia de las matemáticas en las ciencias naturales*, en el que concluía: "que las matemáticas sean el lenguaje apropiado para formular las leyes de la física es un milagro y un don maravilloso que no entendemos ni merecemos".

Es difícil decantarse por una u otra opción. En cualquier caso, estas reflexiones son un buen modo de despedirme de ustedes. En estos siete años de *Juegos Matemáticos,* hemos explorado juntos muchas curiosidades matemáticas o aplicaciones insospechadas al comportamiento humano o social, y también en ocasiones hemos discutido sus limitaciones. Ya sean un mero producto de la mente humana, una realidad más allá de nuestro pensamiento y de la realidad física, o la lógica interna del mundo físico, las matemáticas no dejan de fascinarnos. Todo lo relacionado con las matemáticas que me ha parecido interesante o curioso a lo largo de estos años, y que pudiera explicarse en un lenguaje accesible, ha aparecido en estas páginas. Espero haber despertado en ustedes el mismo interés y fascinación. Les agradezco también todos los comentarios, sugerencias y muestras de aprecio que me han hecho llegar en todo este tiempo, y les pido disculpas si en alguna ocasión no he podido contestar con la profundidad y la atención que merecían.

Quedaba pendiente una pregunta del mes pasado. La solución al "problema lógico más difícil de la historia" en dos preguntas. Sólo tiene solución si el dios de la Confusión elige *a priori* decir la verdad o la mentira (es decir, no elige al azar la respuesta, sino su propio comportamiento al contestar). En este caso, la pregunta (1) al dios A debería ser: "Responderías "ja" a la siguiente pregunta (2): "¿es cierta al menos una de las siguientes afirmaciones: a) vas a responder "da" a la pregunta total (1) y B es el dios de la Verdad, b) B es el dios de la Mentira?"

Supongan, por ejemplo, que "ja" es sí, A es el dios de la Mentira y B el de la Verdad. Si A respondiera "da" a (1), la afirmación a) sería cierta, con lo que la respuesta de A a (2) sería "da" (no). Entonces (1) sería falsa y A, que ha contestado "da" (no), ha dicho la verdad. Si A responde "ja" a (1), ahora a) y b) son ambas falsas, con lo que A respondería "ja" a (2), y habría dicho la verdad al responder "ja" (sí) a (1). O sea, que A no puede mentir, conteste lo que conteste, y, en consecuencia, le estalla la cabeza. Con un poco de paciencia, pueden comprobar que, independientemente del significado de "da" y "ja" y de si A pretende decir la verdad o la mentira, si le explota la cabeza, entonces B es el dios de la Verdad; si la respuesta de A a (1) es "ja", entonces B es el dios de la Mentira; y si es "da", entonces B es el dios de la Confusión.

Por lo tanto, con la pregunta (1) obtenemos la personalidad de B. Si han llegado hasta aquí entendiendo todo este galimatías, no les costará formular la segunda pregunta. Basta utilizar una construcción del tipo "¿responderías "ja" si te preguntara q? La respuesta a esta pregunta, como ya vimos el mes pasado, es "ja" si q es cierta y "da" si q es falsa, independientemente de la personalidad del dios interpelado.