UNIÓN MATEMÁTICA ARGENTINA
FUNDADA EN 1936

NOTICIERO

ISSN 1514-9595 (web)

Misceláneas

El día que Newton se equivocó

Pablo Groisman
Departamento de Matemática
Facultad de Ciencias Exactas y Naturales, UBA
e IMAS-CONICET

Introducción

El 22 de noviembre de 1693, Samuel Pepys le escribió una carta nada menos que a Isaac Newton preguntándole:

¿cuál de las siguientes tres proposiciones tiene más chances de ocurrir?

A. Se lanzan 6 dados equilibrados y obtenemos al menos un ⚅.

B. Se lanzan 12 dados equilibrados y obtenemos al menos dos ⚅.

C. Se lanzan 18 dados equilibrados y obtenemos al menos tres ⚅.

El hecho de que sean 6 lanzamientos (o $2 \cdot 6 = 12$, o $3 \cdot 6 = 18$) y el dado tenga 6 caras no es casual, pero ocuparse de la cara ⚅ no tiene nada de especial para lo que nos ocupa, así que nosotros acá vamos a cambiar el ⚅ por ⚀ para separar la paja del trigo. Nos concentraremos en la cantidad de ⚀ obtenidos en lugar de los ⚅.

Pepys fue un funcionario naval, político y célebre en tanto diarista, pero ejerció como presidente de la Royal Society of London entre 1684 y 1686, justo el momento en que Newton presentó ahí su PhilosophiæNaturalis Principia Mathematica para que fuera publicado luego en 1687. De hecho su nombre aparece en la primera página, autorizando la publicación en nombre de la Real Sociedad. Así que es de esperar que tuvieran cierta relación, como se transluce en la correspondencia. Más aún, parece ser que más que un clásico intercambio epistolar científico, estamos ante un caso de "preguntémosle a mi a amigo matemático que seguro sabe".

Newton le contestó en tres cartas distintas, pero antes de seguir leyendo para encontrarte con su respuesta, te invito disfrazarte de Newton por un rato e intentar una solución para Pepys. ¿Cuál crees que tiene más chances?

El intercambio epistolar puede consultarse en [5]. Ahí Newton le cuenta a Pepys que la probabilidad de A es

$$1 - \mathbb{P}(\text{ningún ⚀ en 6 lanzamientos}) = 1 - \left(\frac{5}{6}\right)^6 = \frac{31031}{46656} = 0.665,$$

la de B es

$$\begin{aligned}&1 - \mathbb{P}(\text{ningún ⚀ en 12 lanzamientos}) - \mathbb{P}(\text{un ⚀ en 12 lanzamientos}) = \\ &1 - \left(\frac{5}{6}\right)^{12} - 12 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{11} = \frac{1346704211}{2176782336} = 0.619,\end{aligned}$$

que es más chica que la de A. Para C Newton dice “En el tercer caso se puede ver que el valor es aún más bajo”. En la carta de Pepys no había opción D, pero si la hubiera habido, podríamos imaginar qué hubiera dicho Sir Isaac: ejercicio para el lector.

Así que la respuesta correcta es la A. Dicho sea de paso, Pepys creía que la respuesta correcta era la C, lo que le valió tener que pagar una apuesta que había hecho con un tal John Smith.

No conforme con la simple respuesta numérica, como tal vez estés vos ahora, Newton intentó buscar una explicación. Si nuestro objetivo es entender y no solo responder correctamente, queremos saber también qué pasa cuando se lanzan 24, 30, 36 y en general $6n$ dados para cualquier n. Y también queremos saber qué pasa si se usan dados con otras cantidades de caras. ¿Es importante que sean justo 6 caras? ¿Qué pasa si son sólo 2 (el caso de las monedas)? Recordar: el objetivo es entender. No generalizar per se. Queremos entender cómo funciona la cosa.

Acá nuevamente les recomiendo parar de leer y jugar un poco con el caso de las monedas. ¿Qué es más probable, al menos una cara en 2 lanzamientos o al menos 2 caras en 4 lanzamientos? ¿y qué de al menos 100 caras en 200 lanzamientos?

Resulta entonces que Newton continuó escribiéndole a Pepys con un argumento bastante discutible, pero que conduce a una respuesta correcta. Newton dice: si dividimos a los dados en grupos de seis, en el caso A necesitamos tener un ⚀ en un grupo, mientras que en B necesitamos un ⚀ en el primer grupo y otro ⚀ en el segundo. Para el C necesitamos un ⚀ en cada grupo. Entonces las probabilidades son cada vez más chicas porque estamos requiriendo más cosas.

El argumento está mal, porque no está considerando la posibilidad de, por ejemplo, sacar dos ⚀ en el primer grupo y ninguno en el segundo. Puede que Sir Isaac haya asumido que podían descartarse esos casos por ser de orden menor, pero no parece ser el caso. Más bien parece que la pifió. Entre otras cosas porque el argumento de Newton no usa el hecho de que los dados son equilibrados. Y adivinen qué. Si los dados no están equilibrados la respuesta A deja de ser la correcta. Por ejemplo si $p =$ probabilidad de que salga ⚀ es $1/4$, entonces B es más probable que A, como pueden chequear ustedes mismos.

Así que el argumento de Newton está mal nomás. No hay mucha vuelta que darle. Stigler [6] se ocupó de recolectar la evidencia. Desde ya, no con el fin de caerle encima a Newton (¡quién podría animarse a semejante cosa!) sino más bien con la muy recomendable idea de aprender de los errores. Además, vale la pena notar que

1. En 1693 la teoría de probabilidades estaba en pañales, dando sus primeros pasos. Los conocimientos del momento eran bastante rudimentarios y ni siquiera había una teoría rigurosa en el sentido matemático (aunque sí había ideas brillantes). Eso llegaría recién 240 años después, en 1933 de la mano de Kolmogorov. La famosa correspondencia entre Fermat y Pascal, que es considerada como el inicio de la

teoría de probabilidades, se desarrolló en 1654. Si veinte años no es nada, cuarenta es dos veces nada, que puede parecer mucho pero es insignificante al lado de los casi 300 años que pasaron hasta el día de hoy. La teoría de probabilidades que tenemos hoy tiene un nivel de desarrollo muy superior al que tenía al momento de la correspondencia entre Pepys y Newton e incluso que el que tenía hace ya casi 100 años, cuando Kolmogorov nos regaló su maravillosa teoría axiomática que metió a la teoría de probabilidades dentro de la casa de la matemática, a la que no pertenecía hasta ese momento.

2. El argumento de Newton lamentablemente es incorrecto, pero hay una idea muy valiosa en él que forma parte central de la teoría moderna de probabilidades. Es lo que hoy conocemos como acoplamientos (coupling). Su desarrollo se dio recién en el siglo XX. Para interiorizarse sobre el asunto puede consultarse el libro de Lindvall [4] o el de Thorisson [7].

Si el argumento de Newton fuera correcto, sería mucho más potente que el cálculo que hicimos arriba porque

1. Es conceptual más que un cálculo. Un cálculo no suele dar razones más allá de que la cuenta da lo que da. Y siempre podemos cometer algún error en la cuenta y llegar a conclusiones erróneas. Un error en un signo, un término que se nos pasa por alto, etc. Un argumento conceptual expone sus falacias -si las tiene- de forma mucho más transparente. Y de hecho eso es lo que ocurre con el argumento de Newton.

2. Un argumento conceptual expone los alcances y limitaciones de la afirmación que estamos intentando demostrar. ¿Es importante que los dados estén equilibrados? ¿Es importante que sean dados de seis caras? ¿Vale sólo para 6, 12 y 18 lanzamientos o para cualquier $6n$ cantidad de lanzamientos? En otras palabras, ayuda a entender.

Acoplamientos

Casi que podríamos decir que los acoplamientos nacieron para poder dar demostraciones conceptuales.

¿Y qué son los acoplamientos? La voz popular dice que son el arte de construir objetos aleatorios en un mismo espacio de probabilidad para poder compararlos.

Uno de los primeros que los puso en práctica en la forma que los conocemos hoy fue Wolfgang Doeblin en los años treinta, para dar una prueba constructiva de la convergencia al equilibrio para cadenas de Markov construyendo cadenas que comienzan con distintas condiciones iniciales en un mismo espacio de probabilidad y mostrando que (gracias a la forma en que las construyó) siempre hay un momento en que se juntan y a partir de ahí siguen juntas para siempre [3]. Eso demuestra que la distribución límite de la cadena no puede depender de las condiciones iniciales y es por lo tanto invariante. Tal es el impacto que produjo que a la condición necesaria para poder hacer esta construcción hoy la llamamos condición de Doeblin. Si tu cadena de Markov la verifica, tu día será más fácil.

Otra gran estrella del arte de acoplar ha sido Anatoliy Skorokhod, que se anotó con dos perlitas: su famoso teorema de representación, que permite intercambiar convergencia en distribución por convergencia casi segura [1, pag. 70] y su teorema de inmersión, que permite meter a cualquier sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con segundo momento finito dentro de un movimiento browniano y obtener así una prueba espectacularmente directa del Teorema Central del Límite [2, Teoremas 37.6 y 37.7].

Casualmente, sobre los métodos directos, Kolmogorov decía

> Es de esperar que ... los probabilistas de la generación más joven, en la locura por el poder de métodos relacionados con las distribuciones en los espacios funcionales, no olviden los métodos directos.

Los acoplamientos son la quintaesencia de los métodos directos. Para entender un poco más de qué se tratan, nada mejor que verlos en acción. Empecemos con un ejemplo simple.

Diremos que la variable aleatoria X tiene distribución de Bernoulli de parámetro p si

$$\mathbb{P}(X=1)=p, \qquad \text{y} \qquad \mathbb{P}(X=0)=1-p,$$

y lo notamos $X\sim$ Be(p). Si $X=X_1+X_2+\cdots+X_n$ con $X_1,\ldots,X_n$ i.i.d con distribución Be(p) decimos que X tiene distribución Binomial de parámetros n y p y lo notamos $X\sim$ Bi(n,p). Es conveniente pensar a las X_i como sucesivas repeticiones independientes de un experimento, siempre en las mismas condiciones y que $X_i=1$ si ocurre determinado resultado y $X_i=0$ si no ocurre. Por ejemplo, podríamos arrojar una moneda cargada con $\mathbb{P}$(cara)$=p$ y entonces $X_i=1$ si sale cara en el i-ésimo lanzamiento y $X_i=0$ si sale ceca.

En los cursos básicos de probabilidad aprendemos que X nos da la cantidad de caras obtenidas en n lanzamientos de la moneda y que

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \qquad k=0,1,2,\ldots,n. \tag{1.1}$$

Creo que todos vamos a coincidir en qué si $X\sim$ Bi(n,p) e $Y\sim$ Bi(n,q) con $p>q$, entonces X es "más grande" que Y en algún sentido (¿en cuál?). Una forma de darle sentido es decir que para todo $k=0,1,2,\ldots,n$,

$$\mathbb{P}(X\geq k)\geq \mathbb{P}(Y\geq k). \tag{1.2}$$

Se la llama dominación estocástica. Si queremos probar (1.2), nuestro primer impulso nos lleva a escribir las expresiones a cada lado del $\geq$ y por lo tanto intentar demostrar que vale la desigualdad

$$\sum_{j=k}^{n}\binom{n}{j}p^j(1-p)^{n-j}\geq \sum_{j=k}^{n}\binom{n}{j}q^j(1-q)^{n-j},$$

para $k=0,1,2,\ldots,n$. (Están invitados a hacerlo). Pero también podríamos proceder de la siguiente manera. Consideremos $U_1,U_2,\ldots,U_n$ variables aleatorias i.i.d. con distribución uniforme en $[0,1]$. Eso significa que para cualquier intervalo $(a,b)\subset[0,1]$, $\mathbb{P}(X\in(a,b))=b-a$. Entonces podemos construir variables $\hat{X}$ e $\hat{Y}$ en un mismo espacio de probabilidad (acopladas) haciendo $\hat{X}_i=1\Leftrightarrow U_i\leq p$, $\hat{Y}_i=1\Leftrightarrow U_i\leq q$ y tomando

$$\hat{X}=\hat{X}_1+\cdots+\hat{X}_n, \qquad \hat{Y}=\hat{Y}_1+\cdots+\hat{Y}_n.$$

Notar que $\hat{X}\sim$ Bi(n,p) e $\hat{Y}\sim$ Bi(n,q). Como $p>q$, por construcción $\hat{Y}_i=1\Rightarrow\hat{X}_i=1$, o lo que es lo mismo, $\hat{X}_i\geq\hat{Y}_i$ y por lo tanto $\hat{X}\geq\hat{Y}$. Entonces el evento $\{\hat{Y}\geq k\}$ está contenido en $\{\hat{X}\geq k\}$ y por lo tanto

$$\mathbb{P}(\hat{X}\geq k)\geq\mathbb{P}(\hat{Y}\geq k).$$

Pero $\mathbb{P}(\hat{X}\geq k)=\mathbb{P}(X\geq k)$ y $\mathbb{P}(\hat{Y}\geq k)=\mathbb{P}(Y\geq k)$, así que

$$\mathbb{P}(X\geq k)\geq\mathbb{P}(Y\geq k).$$

Nos ahorramos la cuenta con las sumas de combinatorios (noten que ni siquiera usamos (1.1)) y de paso expusimos las razones que hacen que X sea “más grande” que Y. Más aún, esta prueba se puede extender a situaciones mucho más generales, mientras que la de las sumas de combinatorios no tanto.

El problema de percolación consiste en considerar un grafo (finito o infinito) y decretar a cada una de sus aristas abierta con probabilidad p o cerrada con probabilidad $1-p$. Todas las elecciones son independientes. Luego fijamos dos nodos del grafo x e y y nos preguntamos por la probabilidad de que exista un camino de aristas abiertas que conecte a x con y. Vamos a denotar a esta probabilidad $\mathbb{P}_p(x \leftrightarrow y)$ para enfatizar que depende de p. Nuevamente, es de esperar que si $p > q$,

$$\mathbb{P}_p(x \leftrightarrow y) > \mathbb{P}_q(x \leftrightarrow y).$$

La demostración con acoplamientos se extiende a este caso casi sin cambios, mientras que con la otra no es claro cómo hacerlo. Involucraría calcular expresiones combinatorias muy complicadas y difícilmente nos lleve a buen puerto.

Resolviendo Pepys-Newton

Volviendo a Newton, trescientos treinta y dos años después, podemos darnos el lujo de tener un argumento que explique la cuestión de fondo más allá de la cuenta y que no deje lugar a dudas. De la mano de los acoplamientos. Y de paso probaremos algo mucho más general que ordenar las probabilidades de A, B y C.

Vamos a considerar dados con cualquier cantidad de caras, que llamaremos $C = 2, 3, 4, \ldots$ y no solo vamos a comparar los casos en que tiramos C, $2C$ o $3C$ dados sino cualquier múltiplo de C.

Teorema

Si se tiran Cn dados equilibrados de C caras cada uno y llamamos p_n a la probabilidad de que salgan al menos n caras ⚀. Entonces $p_1 > p_2 > p_3 > \cdots > p_n$.

Antes de meternos en la demostración, conviene recordar la noción de valor esperado o esperanza de una variable aleatoria. Para una variable aleatoria X, su esperanza se define como

$$\mathbb{E}(X) = \sum_k k\mathbb{P}(X = k),$$

la suma es sobre todos los posibles valores k que puede tomar X. Cuando ese conjunto es infinito, esa suma puede que no esté bien definida, pero acá nos vamos a restringir a variables aleatorias que toman una cantidad finita de valores así que no nos vamos a preocupar por eso.

Para una variable aleatoria $X \sim \mathrm{Bi}(N, p)$ se tiene que $\mathbb{E}(X) = Np$. También vamos a necesitar el siguiente hecho: si Np resulta ser un número entero, entonces $\mathbb{P}(X = k)$ se maximiza (estrictamente) cuando $k = Np$. Pueden probarlo ustedes mismos considerando el cociente $\mathbb{P}(X = k)/\mathbb{P}(X = k-1)$. Llamaremos a este hecho $(*)$. Ahora sí.

Demostración. Seguiremos la buena idea de Newton y acoplaremos las variables que se mencionan en el enunciado del teorema. Vamos a pensar, como Newton, que lanzamos los dados en grupos de C dados. Si llamamos X_n a la cantidad de ⚀ que aparecieron en los primeros n lanzamientos (es decir que se lanzaron Cn dados), tenemos que $p_n = \mathbb{P}(X_n \geq n)$. Las variables aleatorias X_n tienen distribución $\mathrm{Bi}(Cn, 1/C)$, así que su esperanza es $Cn/C = n \in \mathbb{N}$ y por lo tanto para ellas vale $(*)$. Podemos acoplarlas si las construimos haciendo

$$X_{n+1} = X_n + Y,$$

donde Y es la cantidad de dados que mostraron la cara ⚀ en el $(n+1)$–ésimo lanzamiento de C dados.

Newton dijo (o mejor dicho, nuestra interpretación de lo que dijo Newton),

$$p_{n+1} = \mathbb{P}(X_{n+1} \geq n+1) = \mathbb{P}(X_n \geq n, Y \geq 1) = \mathbb{P}(X_n \geq n)\mathbb{P}(Y \geq 1) < \mathbb{P}(X_n \geq n) = p_n,$$

que está mal (¿por qué?), pero con el diario del lunes, podemos arreglarlo. Veremos que $p_{n+1} - p_n < 0$ sin necesidad de calcularlos (gracias a nuestro acoplamiento). En efecto,

$$\begin{aligned}
p_{n+1} - p_n &= \mathbb{P}(X_{n+1} \geq n+1) - \mathbb{P}(X_n \geq n) \\
&= \sum_{r=0}^{C} \mathbb{P}(X_{n+1} \geq n+1 \mid Y = r)\mathbb{P}(Y = r) - \sum_{r=0}^{C} \mathbb{P}(X_n \geq n \mid Y = r)\mathbb{P}(Y = r) \\
&= \sum_{r=0}^{C} \left[\mathbb{P}(X_{n+1} \geq n+1 \mid Y = r) - \mathbb{P}(X_n \geq n \mid Y = r)\right] \mathbb{P}(Y = r) \\
&= \sum_{r=0}^{C} \left[\mathbb{P}(X_{n+1} \geq n+1 \mid Y = r) - \mathbb{P}(X_n \geq n)\right] \mathbb{P}(Y = r) \\
&= \sum_{r=0}^{C} \left[\mathbb{P}(X_n \geq n+1-r) - \mathbb{P}(X_n \geq n)\right] \mathbb{P}(Y = r).
\end{aligned}$$

Para pasar de la primera línea a la segunda usamos la regla de probabilidad total, luego reagrupamos y sacamos factor común $\mathbb{P}(Y = r)$, después usamos que Y es independiente de X_n y por lo tanto las probabilidades referidas a X_n no cambian si condicionamos al valor de Y o si no lo hacemos. Por último usamos el hecho de que si $Y = r$, entonces $X_{n+1} \geq n+1 \Leftrightarrow X_n \geq n+1-r$, y este último evento es independiente de $\{Y = r\}$. Ahora miremos cada uno de los términos entre corchetes en la suma de arriba.

Cuando $r = 0$,

$$\mathbb{P}(X_n \geq n+1) - \mathbb{P}(X_n \geq n) = -\mathbb{P}(X_n = n).$$

Cuando $r = 1$,

$$\mathbb{P}(X_n \geq n+1-1) - \mathbb{P}(X_n \geq n) = 0.$$

Cuando $r = 2$,

$$\mathbb{P}(X_n \geq n+1-2) - \mathbb{P}(X_n \geq n) = \mathbb{P}(X_n = n-1) < \mathbb{P}(X_n = n),$$

gracias a $(*)$. Cuando $r = 3$,

$$\mathbb{P}(X_n \geq n+1-3) - \mathbb{P}(X_n \geq n) = \mathbb{P}(X_n = n-1) + \mathbb{P}(X_n = n-2) < 2\mathbb{P}(X_n = n),$$

nuevamente gracias a $(*)$. Cuando $r = 4$,

$$\mathbb{P}(X_n \geq n+1-4) - \mathbb{P}(X_n \geq n) = \mathbb{P}(X_n = n-1) + \mathbb{P}(X_n = n-2) + \mathbb{P}(X_n = n-3) < 3\mathbb{P}(X_n = n).$$

etc.

Juntando todo tenemos,

$$\begin{aligned}
p_{n+1} - p_n &= \sum_{r=0}^{C} \left[\mathbb{P}(X_n \geq n+1-r) - \mathbb{P}(X_n \geq n)\right] \mathbb{P}(Y = r) \\
&< \sum_{r=0}^{C} (r-1)\mathbb{P}(X_n = n)\mathbb{P}(Y = r) \\
&= \mathbb{P}(X_n = n) \sum_{r=0}^{C} (r-1)\mathbb{P}(Y = r) \\
&= \mathbb{P}(X_n = n) \left(\sum_{r=0}^{C} r\mathbb{P}(Y = r) - \sum_{r=0}^{C} \mathbb{P}(Y = r) \right) \\
&= \mathbb{P}(X_n = n)(\mathbb{E}(Y) - 1).
\end{aligned}$$

Pero $\mathbb{E}(Y) = C \cdot (1/C) = 1$ y por lo tanto

$$p_{n+1} - p_n < 0,$$

que es lo que queríamos probar. □

Así que lo de Newton no estaba taaaan mal. Se le había pasado considerar algunos términos, pero aún teniéndolos en cuenta, la conclusión es la misma. Tal vez simplemente la pifió, o tal vez tenía todo esto en la cabeza y eligió darle una versión simplificada a Pepys. Y el título de este artículo sea sólo un bait para que llegues hasta acá.

Agradecimientos

Quiero agradecer a los integrantes del grupo de probabilidad y modelos estocásticos ModEsto por la lectura crítica y sus valiosos comentarios sobre una versión preliminar del artículo.

Referencias

[1] BILLINGSLEY, Patrick: Convergence of probability measures. John Wiley & Sons, Hoboken, NJ, 1968 (Wiley Ser. Probab. Math. Stat.)

[2] BILLINGSLEY, Patrick: Probability and measure. John Wiley & Sons, Hoboken, NJ, 1979 (Wiley Ser. Probab. Math. Stat.)

[3] DOEBLIN, W.: Éléments d'une théorie générale des chaînes simples constantes de Markoff. In: Ann. Sci. Éc. Norm. Supér. (3) 57 (1940), 61–111. http://dx.doi.org/10.24033/asens.883. – DOI 10.24033/asens.883. – ISSN 0012–9593

[4] LINDVALL, Erik T.: Lectures on the coupling method. New York, NY: Wiley, 1992 (Wiley Ser. Probab. Math. Stat.). – ISBN 0–471–54025–0

[5] RUBIN, Ernest ; SCHELL, Emil D.: Samuel Pepys, Isaac Newton and probability. In: The American Statistician 14 (1960), Nr. 4, S. 27–30

[6] STIGLER, Stephen M.: Isaac Newton as a probabilist. In: Stat. Sci. 21 (2006), Nr. 3, S. 400–403. http://dx.doi.org/10.1214/088342306000000312. – DOI 10.1214/088342306000000312. – ISSN 0883–4237

[7] THORISSON, Hermann: Coupling, stationarity, and regeneration. New York, NY: Springer, 2000 (Probab. Appl.). – ISBN 0–387–98779–7

Pablo Groisman. Es doctor en Ciencias Matemáticas por la Universidad de Buenos Aires e investigador del CONICET. En paralelo con su labor como investigador, Pablo Groisman ha desarrollado una destacada trayectoria como comunicador de la matemática. En el breve período 2022–2026 publicó tres libros —ilustrados por Diego Feld (Gofel)—, lo que pone de manifiesto un compromiso sostenido con la divulgación y el acercamiento de la disciplina a un público amplio.

Su último libro, *La pelota no se ensancha*, acaba de ser publicado y presentado en la Feria Internacional del Libro de Buenos Aires (2026). Aquí compartimos el detalle de las tres publicaciones mencionadas.

- Te regalo un Teorema. Pablo Groisman & Gofel (2022). Editorial Tantaagua.
- Abrazar el azar. Pablo Groisman (2022). Editorial Eudeba.
- La pelota no se ensancha. Pablo Groisman & Gofel (2026). Editorial Tantaagua.