Gregorio R. Moreno Flores

UN PASEO POR EL AZAR

El alucinante universo matemático de las probabilidades: desde los dados hasta el mundo microscópico

Catalonia

GREGORIO R. MORENO FLORES

Es doctor en matemáticas aplicadas y músico de jazz. Inicialmente, planeaba una carrera de músico cuando, al encontrarse con la física, decidió dedicarse a las ciencias. Rápidamente, sus intereses se volcaron hacia las matemáticas, disciplina en la que encontró un mundo de exactitud, belleza e imaginación. Estudió licenciatura y magíster en matemáticas en la Pontificia Universidad Católica de Chile (PUC). Influído por sus mentores, optó por especializarse en la teoría de probabilidades. Obtuvo su doctorado en París. Es profesor de la Facultad de Matemáticas de la PUC. Estudia problemas provenientes de la física, siempre a través del prisma de las probabilidades. Su mundo oscila entre el jazz, los libros (en especial, los cuentos de Borges) y los teoremas.

UN PASEO POR EL AZAR

El alucinante universo matemático de las probabilidades:
desde los dados hasta el mundo microscópico

Gregorio R. Moreno Flores

UN PASEO POR EL AZAR

El alucinante universo matemático de las probabilidades: desde los dados hasta el mundo microscópico

Catalonia

FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación
Gobierno de Chile

Ministerio de Economía, Fomento y Turismo
Gobierno de Chile

milenio
INICIATIVA CIENTÍFICA

NÚCLEO MILENIO
Modelos Estocásticos de Sistemas Complejos y Desordenados

Moreno Flores , Gregorio R.
Un paseo por el azar. El alucinante universo matemático de las probabilidades: desde los dados hasta el mundo microscópico / Gregorio R. Moreno Flores

Santiago de Chile: Catalonia, 2019
148 pp. 15 x 23 cm
ISBN 978-956-324-757-2

PROBABILIDADES & MATEMÁTICAS APLICADAS
519

Diseño de portada y composición: Ximena Morales
Impresión: Salesianos Impresores S.A.
Corrección de textos: Cristine Molina
Dirección editorial: Arturo Infante Reñasco

Primera edición: diciembre 2019
ISBN 978-956-324-757-2

Registro de Propiedad Intelectual N° A-310718

Santa Isabel 1235, Providencia
Santiago de Chile
www.catalonia.cl - @catalonialibros

ÍNDICE

A Alicia.

¿QUÉ ES EL AZAR?

El azar es el motor de una gran cantidad de fenómenos y su entendimiento matemático ha permitido grandes desarrollos humanos. Los modelos matemáticos que intentan explicar los vaivenes del tiempo y de la bolsa de comercio o el comportamiento de sistemas físicos complejos incorporan el azar. Los algoritmos en la tecnología de inteligencia artificial y de minería de datos hacen uso de él. Hoy, se sabe que el azar gobierna el movimiento de los objetos diminutos como las bacterias y los electrones, y juega un papel importante en la teoría de la evolución que explica cómo, a través de escalas de tiempo gigantescas, la genética de los organismos sufrió pequeños cambios aleatorios que convirtieron lentamente una especie en la siguiente.

El entendimiento del azar amplía nuestra capacidad de explicar el mundo y se puede convertir en una valiosa herramienta de análisis. Aquellos que estudiamos el azar demoramos años en dominar sus teoremas más fundamentales y lo cierto es que nunca cesaremos de aprender. Pero un Paseo por el Azar bastará para abrir nuestra mirada a este compañero curioso e impredecible.

¿Dónde está el azar?

Todos nos hemos enfrentado al azar. Estamos sometido a los vaivenes meteorológicos, cada vez más sorprendentes, a la congestión automovilística en los peores momentos y, pareciera, a la acumulación de imprevistos incómodos en los momentos menos adecuados. Por otro lado, también nos encontramos en la calle con ese amigo que no veíamos hace meses o con la información precisa que necesitábamos al inicio de una búsqueda que dábamos por perdida. A veces, compramos el número ganador de una tómbola y llegamos al paradero en el momento preciso en que pasa el autobús de menor frecuencia.

El azar nos trae buenas y malas sorpresas, pero sorpresas, al fin y al cabo. Es particularmente difícil de definir. ¿Quién podría dar en este momento una definición de azar sin rascarse la cabeza por varios minutos y balbucear vanamente una respuesta imprecisa? Más allá de su definición, se presenta ciertamente como un velo nebuloso que cubre todo aquello sobre lo cual no tenemos control. Y, sin embargo, nos fascina de tal modo que hemos desarrollado una serie de mecanismos que hacen uso y abuso de él por mera diversión. Lotería, dados, máquinas tragamonedas, carreras de caballo. El azar parece ser el mayor catalizador de adrenalina.

Nuestros ancestros deben haberse sentido desamparados frente a la antojadiza naturaleza que produce escasez de lluvia o lluvias en exceso, que trae heladas primaverales que dañan los cultivos, que aleja la ruta de las manadas o que, por el contrario, consiente con un invierno corto y clemente y una súbita abundancia de alimentos de fácil cosecha. No es de extrañar que estos fenómenos se atribuyeran a entidades ocultas, todopoderosas, sabias o temperamentales, dependiendo de las culturas.

Por supuesto, hemos progresado en nuestro entendimiento de la naturaleza y en nuestra capacidad de predecir. Hemos erigido, sobre cimientos de roca inquebrantable, un edificio milenario llamado "ciencia" que calma nuestras ansias de explicar el mundo inasible con fórmulas, leyes y teorías. La rigurosa geometría y la intransigente mecánica que describe el movimiento de los objetos son reflejos de una época en la que pensábamos que nuestro conocimiento puede extenderse sin falla desde lo más pequeño hasta la totalidad del universo: midamos el presente con precisión absoluta y sabremos la evolución del futuro. Con el paso del tiempo, hemos renunciado a este afán controlador y le hemos abierto la puerta a la incertidumbre en nuestra descripción del mundo, no por incapaces, sino porque hemos entendido de a poco su rol central.

Hemos descubierto que el azar es una realidad física. El mundo microscópico no se puede describir mediante ecuaciones que entregan plena certeza del estado futuro del sistema estudiado. Hemos observado la naturaleza aleatoria de las mutaciones genéticas que gobiernan el evolucionar de las especies. Hemos vislumbrado el mapa impredecible de los destellos eléctricos de nuestro cerebro que resultan tanto en la mera realización de las tareas cotidianas como en los flujos de idea más impensadas y, también, en la percepción sensorial de nuestro ambiente y nuestra capacidad de analizarlo. También hemos comprendido lo ilusorio que resulta el intentar medir nuestro presente con precisión absoluta y cómo esta limitación también conlleva la pérdida de nuestra capacidad de predecir.

La probabilidad también se presenta como un recurso para lidiar con sistemas extraordinariamente complejos. Pensemos en el tráfico automovilístico en una gran ciudad. Cada habitante aporta su grano de arena a este monstruo impredecible que es la suma de una cantidad astronómica de contribuciones diminutas que se anulan o refuerzan descontroladamente: autos mal estacionados, panas en grandes arterias, peatones imprudentes que demoran el tránsito, congestión en los paraderos o, simplemente, transeúntes corrientes que salen de sus casas en horarios variables. En otro contexto, quien intenta, por ejemplo, entender los niveles de energía de un átomo pesado se encuentra con la sorpresa de que un modelo probabilístico extraordinariamente sencillo puede ser más acertado que una ecuación exacta. Este paradigma —un gran número de perturbaciones ínfimas, individualmente irrelevantes, pero colectivamente ineludibles— se encuentra en numerosos sistemas físicos, biológicos y humanos.

La verdad es que somos francamente malos para lidiar con el azar. Nuestra inteligencia es poco eficiente al estimar las probabilidades de ocurrencia de eventos muy simples y aún más terca a la hora de actualizar nuestras estimaciones iniciales a la luz de nueva información.

Afortunadamente, también hemos sido capaces de entender las leyes del azar. El azar, de algún modo, tiene leyes propias que se pueden someter al análisis matemático. No debe entenderse que este análisis borra la incertidumbre. Nos permite incorporar el azar como un mecanismo poderoso en nuestro modelamiento del mundo y como una fuente de información crucial en la toma de decisiones.

Pese a todos estos avances objetivos, es posible que nuestra percepción cotidiana del azar sea aún parecida a la de nuestros ancestros: el azar como una mano invisible —un *demiurgo*— que gobierna los mecanismos ocultos de lo incierto. ¿No tenemos todos una secreta creencia en la infame *ley de Murphy*? Es posible que estos prejuicios sean la causa del nacimiento tardío de la teoría de probabilidades como disciplina científica.

Nuestra torpeza con las probabilidades nos puede llevar a los razonamientos más absurdos, a conclusiones descabelladas e, incluso, a estrategias de apuesta condenadas al desastre. Las paradojas y los puzles probabilísticos son un tesoro matemático que presentaremos parcialmente en los capítulos que siguen. En varias oportunidades, nos valdremos de ellos para iniciar la discusión de conceptos centrales de la teoría de probabilidades.

Damos por iniciado nuestro paseo por el azar.

Lo probable, lo aleatorio, lo estocástico y el azar: nos encontraremos a menudo con los términos *probable*, *aleatorio*, *estocástico* y *azaroso*. A grandes rasgos, estas cuatro palabras son sinónimas. Cada una de ellas tiene un origen muy antiguo y muy interesante. El origen de las palabras y su evolución a través del tiempo es el campo de una disciplina conocida como ***etimología***.

La palabra ***probabilidad*** llegó a través del francés antiguo y proviene del latín. Designa la verosimilitud de la ocurrencia de algún acontecimiento y se vincula con lo incierto. Originalmente, se refería más a la aceptación de algún hecho como verosímil por parte de los hombres o incluso como demostrable. La palabra adquirió su significado definitivo durante el siglo XVII de la mano con los primeros trabajos matemáticos sobre el azar y, más precisamente, sobre el funcionamiento de los juegos de azar.

La palabra ***aleatorio*** se usa frecuentemente para designar un hecho azaroso. Proviene del latín y se usaba para designar los juegos de dados. Se especula que el origen de esta palabra puede ser el nombre de un hueso, ya que los primeros dados estaban construidos con huesos. Encontramos esta palabra en la famosa frase de Julio César "*Alea iacta est*" ("La suerte está echada"), que pronunció al cruzar con su ejército las aguas del Rubicón, hecho que inició la guerra civil contra Pompeyo.

La palabra ***estocástico*** es un sinónimo de origen griego. Tiene un origen común con la palabra *estaca* y puede guardar referencia con la trayectoria azarosa de un dardo arrojado hacia un blanco.

Finalmente, la palabra ***azar*** desciende del árabe y tiene, quizás, la más bella etimología de entre estos sinónimos. La palabra desciende directamente de la palabra árabe que significa 'dado'. Curiosamente, esta última proviene del nombre de la flor blanca del naranjo y del limón, azahar o flor de azahar. Este vínculo proviene de un dado antiguo con dos caras, una de las cuales se marcaba con el motivo de la flor de naranjo y marcaba la suerte. Hasta hoy en día, la palabra *azar* en árabe se refiere más bien a la suerte y no tanto a la incertidumbre. En inglés, esta palabra adquirió el significado de 'peligro'.

La palabra *probable* tiene cierta connotación positiva. Algo probable tiene cierto grado positivo de verosimilitud. Las palabras *aleatorio*, *estocástico* y *azaroso* son más neutras en este sentido y transmiten un significado más amplio de incertidumbre.

¿Cómo se mide el azar?

¿Qué es medir? Medir consiste en asignar un número a una cantidad física, a un fenómeno o a un atributo de algún objeto. Entre otras cosas, medir nos permite comparar, clasificar y establecer equivalencias. Necesitamos medir distancias para navegar, y medir el peso de un compuesto nos permite asignar la dosis correcta de medicamento a un paciente. También, medir una sutil cantidad de carbono 14 es la clave para determinar la edad de un esqueleto de dinosaurio.

Desde hace milenios, la humanidad ha sido experta en medir todo tipo de cosas. Existe una medida para casi todo lo imaginable: tamaño, peso, volumen, temperatura, velocidad, precio, ángulos, luminosidad, presión, humedad, etc. Más aún, disponemos de instrumentos para medir cada vez con mayor precisión.

Pero ¿se puede medir el azar?

Si bien —desde un punto de vista histórico— hemos sido muy eficientes para medir nuestro entorno, demoramos muchísimo en contestar esta simple pregunta, e incluso en formularla. Los griegos, grandes geómetras, desarrollaron todo tipo de matemáticas, pero no parecen haber llegado a grandes conclusiones respecto a probabilidades. Un obstáculo mayor para esta tarea consistió en la creencia en que lo que percibíamos como azar era una manifestación de alguna entidad oculta y poderosa. Suerte, casualidad, coincidencia, destino son solo algunos de los conceptos que invocamos para referirnos a los hechos inciertos. Curiosamente, los primeros en medir —o cuantificar— el azar fueron apostadores empedernidos en torcerle la mano a la suerte y sacar máximo provecho de los juegos de azar. Ya volveremos sobre este punto.

Para medir un evento azaroso, le asignamos una probabilidad, un número entre 0 y 1, donde 0 significa que el evento no ocurre y 1, que ocurre con la máxima certeza. Es la probabilidad de que el evento ocurra. Cuando pueden ocurrir varias alternativas, le asignamos una probabilidad a cada una de ellas con la regla de que la suma de las probabilidades tiene que ser igual a 1.

A esta altura, el concepto de evento es poco claro, pero se puede entender de manera intuitiva. Básicamente, un evento es alguna circunstancia que podría ocurrir. Por ejemplo, la sentencia "Lloverá mañana" describe un evento. En los informes del tiempo, se asigna una probabilidad a este evento: si se indica que "la probabilidad de precipitaciones es igual a 95%", entonces lloverá con bastante certeza. Por otro lado, si esta probabilidad es igual a 3%, entonces lo más probable es que no llueva.

Hay al menos dos formas de definir la probabilidad. La primera es a través de la frecuencia. Al observar un fenómeno repetidas veces, podemos consignar la frecuencia de veces en las que ocurrió. La frecuencia es el cociente entre el número de veces en que ocurrió y el número total de observaciones. Por ejemplo, si tiramos una moneda 100 veces y salió cara 49 veces, la frecuencia de caras es 49/100, una cantidad muy cercana a ½.

Otra forma de asignar una probabilidad es intentando cuantificar nuestra confianza en que cierto evento ocurrirá, es decir, nuestra apreciación subjetiva de la probabilidad de algún evento. De este modo, se puede dar sentido a oraciones como "La probabilidad de que los senadores voten a favor de cierta ley es 75%". La votación ocurre una sola vez. No es posible repetir la votación un gran número de veces y calcular la frecuencia de votaciones positivas. Sin embargo, con la información de la que disponemos, es posible hacer ciertas predicciones y asignar una probabilidad razonable a este evento. Si nuestra intuición nos lleva a una predicción equivocada, podemos actualizar nuestra apreciación subjetiva integrando esta nueva información.

Porcentajes: en probabilidad, es común expresar cantidades en porcentajes. Decir que tenemos una probabilidad igual a 63% de ganar un cierto juego significa que, si jugamos 100 veces bajo las mismas condiciones, esperamos ganar alrededor de 63 veces; 63% es lo mismo que la fracción 63/100. Ya hablaremos de fracciones con más detalles en la última sección de este capítulo.

Juegos de azar y probabilidades: el nacimiento de una nueva ciencia

La teoría de probabilidades nació de algo tan mundano como intentar entender los juegos de azar. Su desarrollo científico, sin embargo, estimuló un sinfín de creaciones matemáticas que la convirtió en una disciplina poderosa que, a su vez, forma el elemento medular de variadas aplicaciones, desde la física y la biología hasta las finanzas.

El origen de los juegos de azar es antiquísimo. El primero de estos juegos del cual se tiene evidencia sólida es un juego de mesa jugado en Egipto el año 3500 antes de Cristo. El juego, conocido como "perros y chacales", consiste en un tablero por donde los jugadores tienen que hacer transitar unas piezas —los perros y los chacales— cuyos movimientos se determinan al tirar una forma primitiva de dados. Se sabe que los juegos de azar habían llegado a Grecia en 300 antes de Cristo y que en la Roma antigua ya eran tan comunes que se elaboraron leyes para restringir las apuestas.

Todos estos juegos contaban con un artefacto específico para incorporar el azar: el dado. Originalmente, el dado se construía con el hueso de un animal —usualmente, el astrágalo, un hueso del tobillo— cuyas caras se marcaban para diferenciarse. Se sabe de dados tempranos que contaban con cuatro o tan solo dos caras. Con el pasar del tiempo, el dado fue adquiriendo su forma actual. Es interesante notar que el dado y otros artefactos aleatorios también fueron utilizados como herramientas de adivinación.

El *I Ching* es un antiguo libro de adivinación chino conocido también como el Libro de los Cambios. Consiste en 64 hexagramas, figuras formadas por seis líneas horizontales continuas o divididas por la mitad. Cada hexagrama es acompañado de un comentario que se usa en la adivinación. El procedimiento se inicia generando un hexagrama aleatoriamente utilizando varillas, monedas o dados. El hexagrama resultante es entonces analizado por el oráculo basado en el comentario que lo acompaña.

Durante mucho tiempo, los jugadores solo podían valerse de su intuición y experiencia, y no contaban con métodos sólidos para optimizar sus probabilidades de éxito. El concepto de probabilidad como lo entendemos hoy en día demoró siglos en llegar a los diccionarios. En la Grecia antigua, el término *probable* se refería a "aquello que es admitido por todos los hombres o la mayoría o los sabios; todos ellos, la mayoría o los más ilustres". La noción de probabilidad que tenemos hoy en día y su relación con lo incierto parece haberse asentado durante le Edad Media y el Renacimiento.

El primer estudio moderno de las probabilidades del que se sabe hoy en día es un tratado de Girolamo Cardano (1501-1576), escrito alrededor de 1563 y publicado de forma póstuma en 1663. En este trabajo, Cardano estudia juegos de uno, dos y tres dados e introduce la idea de lo equiprobable, es decir, que todas las posibilidades tienen la misma chance de ocurrir. El famoso Galileo (1564-1642) también escribió un pequeño tratado sobre los juegos de dados.

Sin embargo, los verdaderos orígenes de la teoría de probabilidad se encuentran en las inquietudes de Antoine Gombaud (1607-1684), conocido como el Chevalier de Méré, escritor y gran aficionado a los juegos de azar. De Méré se interesó en dos problemas. El primero, conocido como el *problema de los puntos*, consiste en determinar cómo se debe repartir el pozo de un juego que ha sido interrumpido tras algunas jugadas. El segundo consistía en calcular la probabilidad de ganar en cierto juego de dados. Analizaremos ambos problemas en detalle en los capítulos 2 y 4. El problema de los puntos debe probablemente tener un origen anterior a la formulación de De Méré.

En su afán por resolver estos problemas, De Méré contactó al matemático Blaise Pascal, quien inició una extensa correspondencia con su colega Pierre de Fermat. Los matemáticos dieron con una solución a ambos problemas en 1654, dando inicio a un estudio sistemático del cálculo de probabilidades. Estas ideas fueron posteriormente recogidas por Huygens, que publica en 1655 el primer tratado sobre la naciente teoría de probabilidades.

Muchas ideas modernas de esta teoría surgieron durante el siglo XVIII de la mano de los matemáticos Abraham de Moivre, Jacques, Nicolas y Daniel Bernoulli, Euler, Laplace, Lagrange y Fourier, entre otros. Algunos de los teoremas centrales de las probabilidades —como la ley de los números grandes y el teorema central del límite— fueron descubiertos de alguna forma elemental durante esta época. Los siglos siguientes también dieron luz a progresos fundamentales.

Como todas las teorías matemáticas, la teoría de probabilidades se formula en base a axiomas, un conjunto pequeño de conceptos fundamentales sobre los cuales se construyen todos los razonamientos posteriores. En probabilidades, esta formulación fue dada por el matemático Andreï Kolmogorov (1903-1987) en el año 1933 y, a pesar de ser extraordinariamente sencilla, sentó las bases conceptuales de esta teoría.

Ninguna historia de la probabilidad, por muy breve que sea, es completa si no se menciona un objeto central de esta disciplina: el movimiento browniano. En 1828, el botanista Robert Brown observó que pequeñas partículas de polen suspendidas en un líquido oscilan de manera incoherente e impredecible. Brown realizó una serie de experimentos con distintos materiales pulverizados, observando siempre el mismo fenómeno. Llegó a la conclusión de que las partículas que constituyen la materia describen, sin excepción, un rápido movimiento oscilatorio. El fenómeno llama rápidamente la atención de los físicos, quienes realizan experimentos más precisos. En 1905, el físico Albert Einstein establece una primera teoría del movimiento browniano que fue verificada pocos años después con los experimentos de Perrin. En esta época, aún se discutía la naturaleza atómica de la materia. La teoría de Einstein y los experimentos de Perrin pusieron un punto final a esta discusión, concluyendo que toda la materia está formada de diminutas partículas que oscilan "brownianamente". Ambos científicos recibieron el Premio Nobel.

El movimiento browniano fue descubierto de manera independiente y completamente inesperada en un ámbito tan distinto como los mercados financieros. En 1900, el matemático Louis Bachelier publica una tesis sobre las fluctuaciones de los precios de las acciones en la bolsa de comercio. Nuevamente, estas oscilaciones azarosas se pueden describir con el movimiento browniano.

La teoría de probabilidades es un campo de investigación intensamente activo. Las investigaciones teóricas y aplicadas del movimiento browniano y de los otros objetos centrales de la probabilidad florecieron en la segunda mitad del siglo XX y siguen siendo el objeto de un sinnúmero de publicaciones científicas. La teoría de probabilidades es una de las áreas más vivas de las matemáticas y, como toda disciplina científica, congrega una comunidad diversa de investigadores de todos los horizontes. Numerosas revistas científicas solo publican artículos de probabilidades y, cada día, nacen nuevas ideas que fusionan la probabilidad con otras áreas de la matemática y de la ciencia en general.

Todas las disciplinas tienen premios que coronan a sus mejores exponentes. Existen medallas olímpicas, premios Nobel, Óscar y Grammy. Las matemáticas no son la excepción. Si bien no hay un Premio Nobel de matemáticas, existen dos galardones iguales de importantes: el Premio Abel y la Medalla Fields. El segundo tiene la característica de ser entregado solamente a matemáticos de menos de 40 años. Más aún, se entrega solamente cada cuatro años. El matemático Wendelin Werner fue el primer probabilista en recibir la Medalla Fields en 2006. Lo siguieron Stanislav Smirnov en 2010 y Martin Hairer en 2014. El matemático indio Srinivasa Varadhan fue el primer probabilista en recibir el Premio Abel en 2007.

Se atribuye a Isaac Newton la frase "Si he logrado ver más lejos, ha sido porque he subido a hombros de gigantes". Esto es cierto para todos los científicos. Si logramos descubrir ideas nuevas es porque nos hemos alimentado de las ideas de los "gigantes" que nos han precedido y de los "gigantes" contemporáneos a nosotros. Así, el saber matemático se transmite de generación en generación, del profesor al alumno y, en especial, del director de tesis a su estudiante de doctorado. Existe una genealogía de las matemáticas que traza nuestro origen, vinculando a cada matemático con su director de tesis, partiendo de hoy y retrocediendo hasta la Edad Media. Como muchos matemáticos actuales, los orígenes del autor de este libro remontan a los héroes de la probabilidad del siglo XVIII citados anteriormente. El ilustre Srinivasa Varadhan es incluso un descendiente de Newton.

Matemáticos franceses del siglo XVIII: Francia ha sido por siglos la cuna de una de las tradiciones matemáticas más fructíferas de nuestra era. Las bases de la probabilidad como disciplina matemática nacieron de los trabajos de grandes matemáticos franceses del siglo XVIII, como Abraham de Moivre (1667-1754), Joseph Louis Lagrange (1736-1813), Pierre-Simon de Laplace (1749-1827) y Joseph Fourier (1768-1830).

¿Para qué sirven las probabilidades?

Las probabilidades nos dan a lo más una noción de verosimilitud. No nos entregan certeza, a menos que cierta probabilidad sea igual a 0 o a 1, casos en los que, con certeza, podemos asegurar que algo no ocurrirá u ocurrirá respectivamente.

Si hay una probabilidad de 50% de lluvia, no podemos asegurar que lloverá. Sin embargo, este pronóstico no es auspiciador para planificar un picnic al aire libre o un paseo por un parque. Si bien no tenemos certezas, nuestra decisión se verá afectada por este tipo de información y lo más prudente sería elegir un día diferente para nuestra actividad. Una probabilidad de 5% de lluvia quizás no afecte nuestra planificación, puesto que indica una bajísima chance de precipitaciones. El caso opuesto, 95% de probabilidad de lluvia, representa casi una certeza y lo más conveniente será salir con paraguas.

La probabilidad es un tipo de información que puede ser considerada a la hora de tomar una decisión. Saber que algo puede ocurrir nos invita a tomar medidas para anticipar dicho evento aunque no tengamos que recurrir a ellas.

En algunos campos, como ya lo vimos, la incertidumbre es inherente. Los operadores de la bolsa de comercio tienen que tomar decisiones en base a probabilidades, puesto que no tienen certezas sobre el valor futuro de los productos transados. Las compañías de seguros basan su quehacer en nuestro deseo de prevenir hechos negativos como enfermedades, robos, accidentes o incendios. El valor de las pólizas de seguros tiene directa relación con la probabilidad del siniestro al cual están asociadas. En ambos casos, las pérdidas son inevitables. A veces, el precio de una acción baja y, a veces, el cliente cobra un seguro. Lo importante en estas situaciones es que, globalmente, los vaivenes del azar no provoquen pérdidas. La meta es que, en promedio, las transacciones bursátiles o la venta de seguros generen ganancias. Los modelos matemáticos utilizados en ambos contextos incluyen la probabilidad como un elemento esencial.

Algunas consideraciones matemáticas

En este libro, introduciremos los conceptos matemáticos inherentes a la teoría de probabilidad y los utilizaremos con vehemencia. Estos conceptos pueden dar lugar a una discusión matemática profunda que apela a las técnicas más avanzadas de la disciplina; recordemos que la probabilidad es un área de investigación activa que tiene un sitio de honor en la matemática contemporánea. Sin embargo, recurriremos solo a la aritmética elemental en nuestros cálculos, con la excepción notable del capítulo 12, que resultará probablemente más desafiante. El lector familiarizado con fracciones y porcentajes puede omitir los siguientes párrafos.

Hemos indicado que la probabilidad de un evento es un número entre 0 y 1; por lo tanto, la probabilidad no es un número entero, salvo en los casos extremos en los que este valor es exactamente 0 o 1. Los números entre 0 y 1 se expresan mediante su ***expansión decimal***, es decir, un 0 y una coma, seguidos de una sucesión de números enteros. Por ejemplo, 0,24 es un número entre 0 y 1. En el caso del número 0, no es necesario agregar la coma y los dígitos que siguen, puesto que todos ellos son iguales a 0. Es decir, 0 = 0,0 = 0,00 = 0,000 y así. El número 1 se expresa simplemente como 1, aunque el lector recordará la misteriosa identidad 1 = 0,9999… donde el 0 y la coma son sucedidos por infinitos 9. Recordemos también que agregar ceros a la derecha no altera el valor de una expansión decimal; por ejemplo, 0,5 = 0,50 = 0,500. En general, todos los números, sin importar su tamaño, tienen una expansión decimal. Así, 34,68 es un número entre 34 y 35. Usaremos la coma para separar la parte entera de un número del resto de su expansión decimal. Es muy común sustituir la coma por un punto, sobre todo en los computadores y en las famosas calculadoras, cada vez más en desuso (excepto por la calculadora del celular, siempre tan útil a la hora de dividir una cuenta en un restaurant). Reservaremos el uso del punto para denotar números grandes, agrupando sus dígitos de tres en tres, por ejemplo, escribiendo

2.302.840 en lugar de 2302840. Esto facilita la lectura de los números a veces gigantescos que aparecerán en nuestros cálculos.

En la gran mayoría de los ejemplos de este libro, los resultados serán dados por el cociente de dos números enteros, esto es, una ***fracción***. A pesar de que las fracciones son la pesadilla de muchos estudiantes, su uso es bastante sencillo si recordamos algunas leyes elementales (ver el recuadro más adelante). Algunas expansiones decimales se pueden escribir fácilmente como una fracción. Por ejemplo, 0,5 = ½ o 0,25 = ¼; estos valores aparecerán frecuentemente en nuestros cálculos. Algunas fracciones resultan en expansiones decimales largas —a veces infinitas—, pero nos restringiremos a escribir dos decimales. Cada vez que recurramos a estas "truncaciones", usaremos el símbolo ≈ para indicar que se trata de una aproximación. Por ejemplo, nos encontraremos frecuentemente con los valores ⅓ ≈ 0,33 y 1/6 ≈ 0,17. Estas fracciones tienen una expansión decimal infinita, por lo que solo las aproximaremos. Recordemos también que algunos números no son fracciones, como $\pi \approx 3{,}14$.

Las fracciones tienen ***numerador*** y ***denominador***: el número de arriba y el de abajo. Las fracciones cuyo denominador es 100 tienen una expansión decimal muy sencilla. Por ejemplo, 37/100 = 0,37. También, 0,25 = 25/100 y, recordando que los ceros a la derecha no alteran el valor del número, 0,5 = 0,50 = 50/100. Notemos que ya habíamos escrito estos valores como fracciones distintas: ¼ y ½ respectivamente. Esto agrega desgraciadamente una capa de dificultad adicional: las fracciones se pueden expresar de más de una manera. Sin embargo, no será un problema mayor en nuestros cálculos. Recordemos otro hecho importante: cuando el numerador y el denominador son iguales, el resultado es simplemente igual a 1. Por ejemplo, 2/2 = 1 y 4/4=1.

Las fracciones cuyo denominador es 100 tienen una representación natural como ***porcentajes***. Volviendo al ejemplo de más arriba, 0,37 = 37%. Esto nos da tres representaciones posibles de un número: como expansión decimal, como fracción y como porcentaje. Por sobre todo, tenemos que recordar las dos identidades siguientes: ½ = 0,5 = 50% y ¼ = 0,25 = 25%. Ambas aparecerán frecuentemente. También será necesario recordar que 0% = 0 y 100% = 1.

Finalmente, anticipamos que la probabilidad de un evento puede ser extremadamente pequeña. Nuestro método de redondear a dos decimales no nos permitirá entonces discernir entre 0,5412 y 0,5447, o entre 0 y 0,0004. En los casos excepcionales en los que esta distinción es necesaria, usaremos porcentajes con dos decimales. Por ejemplo, 0,5412 = 54,12%, 0,5447 = 54,47% y 0,0004 = 0,04%. A través de estos ejemplos, el lector podrá darse cuenta de que estas conversiones son bastante sencillas. Estas últimas consideraciones solo se usarán en el capítulo 12, cuya complejidad matemática es un poco mayor.

Existen varios sitios web gratuitos que permiten realizar cálculos en algunos segundos. Se recomienda al lector utilizar el sitio *wolframalpha.com*, en especial para los cálculos que involucran números extremadamente grandes o extremadamente pequeños.

Suma y producto de fracciones: en los ejemplos de este libro será a menudo necesario operar con fracciones. El producto de fracciones es sencillo. Basta con multiplicar los numeradores entre sí y los denominadores entre sí. Por ejemplo, 3/5 × 2/6 = 6/30. La suma de fracciones es un poco más complicada. Cuando dos fracciones tienen el mismo denominador, basta con sumar los numeradores: 2/7 + 3/7 = 5/7. Cuando los denominadores son distintos, la situación se vuelve más compleja: tenemos que escribir las fracciones con el mismo denominador. Afortunadamente, hay pocos casos en los que esto será necesario en este libro. En particular, nos encontraremos con el cálculo ¼ + ½ = ¾. Esto se obtiene notando que ½ = 2/4. Por lo tanto, ¼ + ½ = ¼ + 2/4 = ¾. Las fracciones también se pueden simplificar dividiendo numerador y denominador por un mismo número. Por ejemplo, 3/6 = ½.

Nuestro paseo por el azar

La organización de este libro es más o menos así:

Iniciamos nuestra discusión en el capítulo 2 con el ejemplo más sencillo de todos: el juego de cara o sello. Esto nos permitirá introducir ideas complejas de forma elemental.

El capítulo 3 contiene las nociones medulares de la teoría de probabilidades. En él se definen los principales conceptos que se usarán en el resto del libro. El capítulo 4 contiene aplicaciones de estas ideas. Entre otras cosas, se contestan las preguntas del Chevalier de Méré.

El capítulo 5 desarrolla la idea de distribución de probabilidad insinuada en el capítulo 3.

Los capítulos 6 y 7 nos presentan una serie de paradojas intrigantes en torno al concepto de probabilidad condicional.

El capítulo 8 termina nuestra presentación de los elementos teóricos de probabilidades al discutir las variables aleatorias y su promedio.

El capítulo 9 es un sobrevuelo de la teoría de los procesos estocásticos, los objetos probabilísticos que describen el movimiento, con un énfasis en su historia y aplicaciones. El capítulo 10 describe algunas aplicaciones de estas ideas al funcionamiento de los motores de búsqueda y de las redes sociales. El capítulo 11 discute la historia del proceso estocástico más famoso de la teoría: el movimiento browniano.

El capítulo 12 nos entrega herramientas matemáticas más avanzadas para el cálculo de probabilidades. Es sin duda el capítulo más complejo desde el punto de vista matemático y se puede omitir sin perjuicio de la comprensión del resto del libro.

Finalmente, el capítulo 13 discute algunos juegos probabilísticos cuyas soluciones están fuera del alcance de las técnicas desarrolladas en los capítulos anteriores. Sin embargo, el planteamiento de los problemas es accesible y la discusión de las soluciones es motivo para introducir algunas ideas nuevas. El último problema es elemental y se deja como un desafío para el lector.

¿CARA O SELLO?

PROBABILIDAD ½

El juego de cara o sello es sin duda el más simple y más popular juego de azar. Cuando queremos echar alguna decisión a la suerte, podemos tirar una moneda y tomar dicha decisión basados en el resultado: si sale cara, hago tal cosa y, si sale sello, no lo hago. También el cara o sello es el instrumento oficial para decidir cuál de los dos equipos inicia un partido de fútbol.

Básicamente, tiramos una moneda cuando estamos indecisos y cuando escoger una u otra alternativa es más o menos irrelevante. Tirar una moneda nos asegura que ambas alternativas tienen la misma chance de ser escogidas: ambas serán escogidas con probabilidad ½.

Sin embargo, esto no significa que, si tiramos una moneda una cierta cantidad de veces, exactamente la mitad de las veces mostrará cara y la otra mitad, sello. Si lanzamos una moneda cinco veces, puede perfectamente salir cara cuatro veces y sello una sola vez. Sin embargo, si la lanzamos 100 veces, lo más probable es que la cantidad de caras y de sellos sea cercana a 50; algo así como 47 y 53 o 51 y 49. Es decir, la frecuencia de caras y de sellos es muy cercana a ½ cuando repetimos el juego una gran cantidad de veces. Esto refleja el hecho de que la probabilidad de cara y de sello son ambas iguales a ½.

Una nota de advertencia: existen máquinas que lanzan monedas de modo de obtener exactamente el resultado deseado. Algunos magos entrenados pueden lograr lo mismo controlando precisamente la fuerza con la que arrojan la moneda. Sin embargo, lanzando la moneda enérgicamente al aire, el resultado es, para todos los efectos, completamente aleatorio.

Juegos repetidos

Si tiramos una moneda, pueden ocurrir dos posibilidades: cara o sello, ambas con probabilidad ½. Si tiramos dos monedas, las posibilidades aumentan: cara-cara, cara-sello, sello-cara y sello-sello. Todas estas posibilidades debieran ocurrir con la misma probabilidad y, como la suma total de las probabilidades es igual a 1, esta probabilidad debe ser ¼.

Esto nos permite calcular probabilidades más complicadas. Al tirar dos monedas, ¿cuál es la probabilidad de que ambas muestren el mismo resultado? Solo tenemos que contar las posibilidades favorables: cara-cara y sello-sello. Luego, la probabilidad buscada es la probabilidad de que ocurra doble cara o doble sello, o sea, la suma de ambas probabilidades: ¼ + ¼ = ½.

Ahora, con tres monedas. La siguiente tabla muestra las distintas posibilidades:

Primera moneda	Segunda moneda	Tercera moneda
Cara	Cara	Cara
Cara	Cara	Sello
Cara	Sello	Cara
Cara	Sello	Sello
Sello	Cara	Cara
Sello	Cara	Sello
Sello	Sello	Cara
Sello	Sello	Sello

Esta vez, hay ocho posibilidades. Todas ellas debieran tener la misma probabilidad. Como estas probabilidades tienen que sumar 1, todas ellas deben ser iguales a ⅛ ≈ 0,12.

¿Cuál es la probabilidad de que las tres monedas muestren el mismo resultado? Hay dos posibilidades favorables: la primera y la última. La probabilidad buscada es, por lo tanto, ⅛ + ⅛ = 2/8 = ¼ = 0,25.

¿Cuál es la probabilidad de obtener exactamente dos caras? Las posibilidades favorables son la segunda, tercera y quinta. Luego, la probabilidad buscada es ⅛ + ⅛ + ⅛ = ⅜ ≈ 0,37.

Otra forma de calcular probabilidad, como lo veremos más adelante, consiste en calcular el cociente entre las posibilidades favorables y las posibilidades totales. En el último cálculo, hay 3 posibilidades favorables y 8 posibilidades totales. Por lo tanto, la probabilidad buscada es $3/8$, tal como ya sabíamos.

Cuando aumentamos la cantidad de monedas, aumenta dramáticamente la cantidad de posibilidades y los cálculos se vuelven cada vez más complicados. Sin embargo, observamos un patrón sencillo: para una moneda, hay dos posibilidades; para dos monedas, cuatro posibilidades y, para tres monedas, ocho posibilidades. ¿Qué podemos inferir para cuatro o más monedas?

Nos encontramos con el **carácter multiplicativo** del conteo de posibilidades. Para cada moneda que lancemos, hay dos posibilidades. Luego, si tiramos varias monedas, el número total de posibilidades será una **potencia de 2**, es decir, 2 multiplicado consigo mismo un cierto número de veces. Para una moneda, obtenemos $2^1 = 2$ posibilidades. Para dos monedas, las posibilidades suben a $2 \times 2 = 2^2 = 4$, esto es, 2 multiplicado consigo mismo dos veces, también conocido como 2 al cuadrado. Para tres monedas, $2 \times 2 \times 2 = 2^3 = 8$ posibilidades, o sea, 2 al cubo. Para cuatro monedas, el número total de posibilidades debe ser, por lo tanto, igual a $2 \times 2 \times 2 \times 2 = 2^4 = 16$ - esto es, 2 a la potencia 4 o 2 elevado a 4. Y así sucesivamente.

Potencias de 2 y computadores: las sucesivas potencias de 2 del 1 al 10 son 2, 4, 8, 16, 32, 64, 128, 256, 512 y 1.024. El lector probablemente reconocerá estos números, especialmente los más grandes, como cantidades frecuentes en computación. Esto se debe a que los computadores están basados en un sistema binario. A nivel elemental, la información se codifica como secuencias de ceros y unos. Un bit —el acrónimo de *binary digit*— puede almacenar el valor 0 o 1. Un byte corresponde a 8 bits, mientras que un kibibite —1 KiB— consta de 1.024 bytes. El kibibite no debe confundirse con el kilobyte = 1.000 bytes, aunque sean números muy similares. Esta distinción ha sido fuente de confusiones.

Esta "multiplicatividad" se traduce en un carácter multiplicativo de las probabilidades. Antes de entrar en detalles, vamos a introducir una notación muy útil. Vamos a representar las caras por C y los sellos por S. Así, la secuencia CSSC corresponde a cuatro tiradas consecutivas de una moneda cuyo resultado fue cara, sello, sello y cara.

¿Cuál es la probabilidad de obtener la secuencia CSSC al tirar cuatro veces una moneda? Basado en nuestro argumento anterior, hay $2^4 = 16$ posibilidades que tienen cada una la misma probabilidad de ocurrir. Luego, la probabilidad de obtener esta secuencia es igual a $1/16 \approx 0{,}06$.

Pensemos puramente en términos de probabilidades. La probabilidad de obtener C en la primera tirada es ½. La probabilidad de obtener S en la segunda es también ½, y así con la tercera y la cuarta tirada. Si multiplicamos estas probabilidades, obtenemos $½ \times ½ \times ½ \times ½ = 1/16$, el mismo resultado anterior. Las probabilidades se multiplican.

Esta multiplicatividad de las probabilidades no es un hecho general. Proviene en este caso de la **independencia** de las sucesivas tiradas de la moneda, es decir, el resultado de la primera tirada no tiene incidencia sobre el resultado de las siguientes, y viceversa. Volveremos a discutir la independencia en el próximo capítulo.

Secuencias aleatorias, ADN y la falacia del jugador

No es fácil fingir una secuencia aleatoria de caras y sellos. Considere las dos secuencias siguientes:

1.- CCCCSCCSCSSSSSSCSSCSSSSSCS
2.- CSSCSCSCCSCCSSSCCSCCCSCSS

Una de ellas es aleatoria y la otra, inventada. Cuando intentamos emular una secuencia aleatoria, tenemos tendencia a balancear las C y las S. Pensamos que repeticiones de C se deben compensar con repeticiones de S o que no puede haber una acumulación de resultados de uno de los dos tipos. Esto es falso. En una secuencia larga, aparecerán repeticiones largas de C y de S. Después de todo, las secuencias CCCCCC y CSCSCS tienen la misma probabilidad de aparecer. Es más, en una secuencia infinita de cara o sello, cualquier patrón se repetirá infinitas veces y habrá sucesiones de C y de S de largo arbitrariamente grande. ¿Puede el lector adivinar cuál es la secuencia aleatoria en el ejemplo anterior?

Esto tiene que ver con la infame **falacia del jugador**. Pensar que una larga secuencia de C debe compensarse pronto con la aparición de una S equivale a decir que las jugadas pasadas afectan las jugadas futuras, argumento que viola la noción de independencia. Muchos apostadores perdieron una cantidad considerable de dinero en un juego de ruleta el 18 de agosto de 1913 en el casino de Montecarlo. En la ruleta, se puede apostar al color, rojo o negro, los cuales aparecen con probabilidad ½ cada uno. Esa noche, la ruleta marcó el color negro 26 veces seguidas. Después de numerosas apariciones del color negro, muchos jugadores apostaron repetidamente al color rojo, perdiendo una gran cantidad de dinero en el proceso. En la jugada número 27, la aguja de la ruleta apuntó, por fin, al color rojo.

La martingala es una estrategia de apuesta en desuso que consiste, tras cada pérdida, en apostar el doble de lo perdido.

Consideremos el caso de un jugador de cara o sello. El jugador apuesta inicialmente un doblón. Si gana, se lleva un premio de un doblón. Si pierde, apuesta dos doblones en la siguiente jugada. Si gana esta vez, gana dos doblones, compensa su pérdida inicial y se queda, de hecho, con un beneficio neto de un doblón. Por el contrario, si pierde, vuelve a apostar cuatro doblones y así sucesivamente. Se puede verificar que, de este modo, la primera vez que gana, recupera lo perdido y un doblón adicional. A primera vista, pareciera que esta estrategia lleva a una ganancia segura. El problema es que este esquema de apuesta es extremadamente caro y lleva rápidamente a la bancarrota.

El origen del nombre *martingala* se debe al gentilicio de los habitantes de Martigues, una región aislada de Francia, conocidos por su ingenuidad. Así, jugar *à la martingala* se refería a jugar de manera completamente absurda.

Hoy en día, el término *martingala* se aplica a una familia de objetos centrales de la teoría de probabilidades que guardan poca relación con la estrategia de apuestas.

Estimar el largo de la mayor secuencia de caras o de sellos en una cierta cantidad de jugadas es un problema clásico de probabilidades. Se sabe que, en promedio, es aproximadamente igual al logaritmo del número total de jugadas dividido por el logaritmo de 2. Así, en 1.000.000 de jugadas, es esperable ver una secuencia de unas 20 caras o 20 sellos sucesivos. En 1.000.000.000 de jugadas, este número aumenta a 30, mientras que, en 1.000 jugadas, es del orden de 10.

El problema de las "secuencias largas" tiene aplicación en el estudio de secuencias de ADN. Al comparar dos secuencias de ADN, importa determinar el largo de los segmentos en los que coinciden. Matemáticamente, este problema es una generalización del caso considerado en el párrafo anterior.

Resolviendo el problema de los puntos

Recordemos el ***problema de los puntos*** planteado por el Caballero de Méré en el siglo XVII. Dos personas, A y B, juegan al cara o sello y contabilizan la cantidad de caras y de sellos en tiradas sucesivas. Si salen primero tres caras (no necesariamente contiguas), gana A; si salen primero tres sellos, gana B. El juego se detiene cuando ocurre alguna de estas dos posibilidades. Por ejemplo, si la sucesión de caras y sellos es cara-cara-sello-sello-cara, gana A; pero si la sucesión es cara-sello-cara-sello-sello, gana B. Ambos apuestan la misma cantidad de dinero. El ganador se lleva el total.

¿Cuánto puede durar el juego? No mucho, la verdad. Después de cinco tiradas, al menos una de las dos caras de la moneda debe haber aparecido al menos tres veces. Podemos convencernos de este hecho listando todas las posibilidades: cinco caras; cuatro caras y un sello; tres caras y dos sellos; dos caras y tres sellos; una cara y cuatro sellos; cinco sellos. En los tres primeros casos, hay al menos tres caras, mientras que, en los tres últimos casos, hay al menos tres sellos. Observamos también que ocurre lo uno o lo otro: si hay al menos tres caras, hay a lo más dos sellos y viceversa. La cuenta anterior nos permite también concluir que la probabilidad de que gane cualquiera de los dos jugadores es ½. Es decir, el juego es justo.

Reducción al absurdo: podemos también llegar a esta conclusión por un clásico argumento por contradicción o “reducción al absurdo”. Supongamos que ninguna de las dos posibilidades, cara y sello, ocurrió más de dos veces en cinco jugadas. Luego, hay a lo más dos tiradas con resultado cara y a lo más dos tiradas con resultado sello. Sumando, vemos que hay a lo más 2 + 2 = 4 tiradas en total. Como sabemos que el número total de tiradas es 5, esto es una contradicción. En otras palabras, la conclusión es “absurda”. Por lo tanto, alguna de nuestras premisas debe ser falsa. La única suposición dudosa que hicimos fue que ninguna de las dos posibilidades ocurre más de dos veces. Esto debe, por lo tanto, ser falso. Conclusión: alguna de las dos posibilidades ocurre al menos tres veces en cinco tiradas.

Supongamos ahora que, por alguna razón misteriosa, el juego se detiene a la primera jugada. ¿Cómo dividimos el pozo? Si sale cara, A lleva ciertamente la delantera. ¿Tiene que llevarse A la totalidad del pozo? ¡Claro que no! Las siguientes tres jugadas podrían perfectamente arrojar tres sellos y, por lo tanto, ganaría B. De hecho, hay varias otras posibilidades en las que gana B. B tiene que llevarse una parte del pozo, aunque sea menor.

Una posibilidad más elaborada consiste en dividir el poso de acuerdo con la proporción de caras y sellos obtenida hasta la interrupción del juego. De este modo, si el juego se detiene a la segunda jugada y las tiradas sucesivas dieron cara y sello, el pozo se divide por la mitad. Esto parece justo. Sin embargo, si salieron dos caras, la proporción de cara es 1 y la proporción de sellos es 0. Por lo tanto, según esta política, A se lleva todo el pozo. Sin embargo, B aún tiene chances de ganar, aunque sean muy pequeñas.

La solución propuesta por Pascal es mucho más justa: el pozo debe dividirse de acuerdo con las probabilidades de ganar de cada jugador. Demos algunos ejemplos:

Si el juego se interrumpe después de dos jugadas y el resultado fue una cara y un sello, es claro que nadie lleva la delantera. Aun sin contar posibilidades, no hay ninguna duda de que ambos jugadores tienen una probabilidad ½ de ganar. El pozo debe, por lo tanto, dividirse por la mitad.

Ahora, si el juego se interrumpe después de dos jugadas y el resultado fue dos caras, la única posibilidad de que gane B es que salgan tres sellos en las siguientes tres jugadas. Como calculamos anteriormente, este evento tiene probabilidad ⅛. B debe, por lo tanto, llevarse un octavo del pozo y A, siete octavos.

Estudiemos ahora una situación más elaborada. Supongamos que el juego se detiene a la tercera jugada y que el marcador indica dos caras a favor de A y tan solo un sello a favor de B. ¿Cómo debe repartirse el pozo? En otras palabras, ¿cuál es la probabilidad de que gane cada jugador? La única posibilidad para que gane B es que las dos tiradas siguientes arrojen sello. Como ya sabemos, esto ocurre con probabilidad ¼. Por lo tanto, B se lleva un cuarto del pozo, mientras que A se lleva los tres cuartos restantes.

La ley de los números grandes

Si lanzamos una moneda al aire, las probabilidades de obtener cara y sello son ambas iguales a ½. Sin embargo, esto de ninguna manera significa que, cuando arrojamos una moneda repetidas veces, obtenemos cara exactamente la mitad de las veces y sello la otra mitad. De hecho, si lanzamos una moneda una cantidad impar de veces, esta situación es imposible. Intentemos convencernos de que esto es falso aun si lanzamos la moneda una cantidad par de veces.

Paradójicamente, si lanzamos la moneda solo dos veces, ¡esto resulta ser cierto! Calculemos la probabilidad de obtener una cara y un sello al lanzar dos monedas. Las posibilidades favorables son cara-sello y sello-cara. Ya sabemos que ambas tienen una probabilidad igual a ¼ de ocurrir y, por lo tanto, la probabilidad de obtener una cara y un sello en dos tiradas es ¼ + ¼ = ½.

Estudiemos entonces lo que ocurre con cuatro tiradas; siendo 3 un número impar, este caso no es relevante. Sabemos que hay 16 posibilidades en total. Las posibilidades que arrojan exactamente dos caras y dos sellos son seis: CCSS, CSCS, CSSC, SSCC, SCSC, SCCS. Como hay 16 posibilidades distintas para el resultado de cuatro monedas, cada una de las seis configuraciones anteriores tiene una probabilidad igual a 1/16 de ocurrir. Sumando estas probabilidades, vemos que la probabilidad de obtener exactamente dos caras y dos sellos en cuatro tiradas es $6/16 \approx 0{,}37$, una cantidad menor a ½; de hecho, ½ = 8/16.

Lo que ocurre realmente es que la proporción de caras se acerca a ½ cuando tiramos la moneda una cantidad muy grande de veces, y lo mismo para la proporción de sellos. Esta es la ***ley de los números grandes***. Esta importante ley de la teoría de probabilidades asegura lo siguiente: si, en un experimento, un evento tiene una probabilidad igual a **p** de ocurrir, entonces, al repetir el experimento una gran cantidad de veces, la proporción de veces en que el evento ocurrió es muy cercana a **p**. Esta proporción quizás no sea nunca igual a **p**, pero se volverá muy cercana.

Podemos preguntarnos qué significa realmente que la probabilidad de cara sea igual a ½. Es un supuesto bastante razonable y expresa nuestra expectativa de que la moneda es físicamente balanceada. Si al lanzar la moneda una gran cantidad de veces vemos que la proporción de caras es mucho menor, tendremos que actualizar nuestro supuesto. ¿Será esto posible? Conversaremos de monedas y dados "alterados" en el capítulo 3.

La ley de los números grandes y la falacia del jugador: pareciera que la ley de los números grandes argumenta a favor de la falacia del jugador. Sin embargo, como su nombre lo indica, solo es aplicable a un número grande de jugadas. En una cantidad pequeña de jugadas, puede pasar virtualmente cualquier cosa. Los caras y sellos se balancean solo a largo plazo. Aun así, la cantidad de caras podría ser mucho mayor que la cantidad de sellos, o viceversa. Lo importante es que esa diferencia es pequeña comparada con el número total de jugadas.

No todo es equitativo: el cara o sello es el mejor representante de una situación paradigmática en la que exactamente dos posibilidades pueden ocurrir, ambas con probabilidad ½. Sin embargo, esto no es siempre el caso. Se sabe, por ejemplo, que la probabilidad de que nazca una niña es levemente mayor que la probabilidad de que nazca un niño. Esto se refleja en el simple hecho de que la proporción de mujeres en el mundo es levemente mayor a la proporción de hombres. Asimismo, la probabilidad de lluvia en un día dado es rara vez igual a ½ y depende de manera complicada de una gran cantidad de variables. También veremos más adelante que, en un cierto juego de dados, la probabilidad de ganar es levemente mayor a ½. Esto, sin embargo, es atípico: en los juegos de azar, la casa tiene usualmente las probabilidades a su favor.

EL ESPACIO DE PROBABILIDAD

¿Qué tienen en común el cara o sello, los juegos de dados, el póker y la lotería? ¿Cuáles son las herramientas matemáticas que nos permiten describir los fenómenos aleatorios de forma sencilla?

En este capítulo, conoceremos el marco teórico que permite formalizar toda la probabilidad: el espacio de probabilidad.

Monedas, dados, cartas y loterías

Para calcular probabilidades, lo primero consiste en listar todas las posibilidades. Luego, le asignamos a cada posibilidad una probabilidad: un número entre 0 y 1 para cada posibilidad, de modo que todos ellos sumen exactamente 1. Los juegos de monedas, dados, cartas y lotería nos dan ejemplos de situaciones en las que todas las posibilidades tienen la misma probabilidad, es decir, son **equiprobables**. Cara y sello tienen la misma probabilidad de ocurrir. Los resultados del dado aparecen todos con la misma frecuencia. Al sortear una carta al azar, ninguna tiene mayores chances de ser elegida que las demás. No hay números de la suerte en la lotería.

Ya vimos que en el juego de cara o sello hay solo dos posibilidades. Necesariamente, ambas tienen probabilidad ½. Cuando arrojamos un dado, hay seis posibilidades: los números del 1 al 6. Como todas ellas tienen la misma chance de aparecer, su probabilidad debe ser ⅙ ≈ 0,17. En esencia, un dado es una moneda con seis caras distintas o una moneda es simplemente un dado con dos caras. Notemos que, hasta aquí, la equiprobabilidad es una verdad teórica: confiamos en que ninguna cara de la moneda o del dado es privilegiada. Como en el cara o sello, la ley de los números grandes nos permite confrontar esta creencia con la realidad.

Un mazo de cartas está formado de 52 cartas distintas, al menos en el caso de famosa baraja inglesa, que es una de las más populares. Al seleccionar una carta al azar, la probabilidad de que salga cualquiera de ellas es por lo tanto igual a 1/52. Por ejemplo, si revolvemos las cartas y escogemos una al azar, la probabilidad de que la carta escogida sea el 6 de corazones es 1/52. Lo mismo para el 7 de espadas y para cualquier otra carta. Una mano de póker consiste en cinco cartas. Todas las posibles combinaciones de cinco cartas tienen la misma probabilidad. El problema es que es más difícil determinar el número de casos totales. La respuesta es 2.598.960. Ya veremos este cálculo con más detalle en el capítulo 12.

En el loto, el jugador tiene que escoger seis números entre el 1 y el 41. El día del sorteo, una máquina elige seis números al azar. Para ganar, se debe acertar los números sorteados. Aquí no hay trucos: cualquier combinación de seis números tiene la misma probabilidad de ser escogida por la máquina y hay 4.496.388 de ellas. La probabilidad de acertar es 1/4.496.388, un número inimaginablemente pequeño.

Hay un dato curioso acerca de la lotería: los jugadores prefieren ciertos números, como el 7 o el 13. También hay una tendencia a elegir números dispersos o a evitar números en el borde de la ficha. El sorteo es ciego ante estas preferencias. Cuando los números ganadores coinciden con estos números "favoritos", el pozo debe dividirse entre una mayor cantidad de ganadores. Por lo tanto, es aconsejable elegir un patrón de números atípicos. Así, en el improbable caso de ganar, el pozo se repartirá entre menos jugadores y las ganancias serán bastante mayores.

El espacio de probabilidad y la probabilidad uniforme

Estableceremos las bases conceptuales de la teoría de probabilidades. Esta sección es quizás la más abstracta de este libro. Sin embargo, dar definiciones precisas de todos los conceptos que utilizaremos nos permite disponer de un vocabulario fluido para describir nuestro trabajo futuro. Después de todo, la probabilidad es una teoría matemática y, como tal, se debe erigir sobre cimientos rigurosos.

Una situación hipotética en la que opera el azar suele llamarse un **experimento**. Por ejemplo, tirar una moneda es un experimento. Lanzar un dado es un experimento, al igual que escoger al azar una bola de lotería. Esto no coincide con el concepto habitual de experimento —incluido el laboratorio y la maquinaria sofisticada—, pero es la terminología estándar de las probabilidades y de la estadística.

El ***espacio muestral*** consiste en todos los resultados posibles del experimento, es decir, todas las posibilidades. Estas posibilidades también se conocen como **elementos del espacio muestral**. Por ejemplo, en el experimento "tirar una moneda", el espacio muestral consiste en las posibilidades cara y sello. En el experimento "lanzar un dado", el espacio muestral corresponde a los números enteros entre el 1 y el 6. Un ***evento*** es un conjunto de elementos del espacio muestral, es decir, una lista de posibles resultados del experimento. Si el resultado del experimento cae dentro de este conjunto, decimos que el evento ocurrió. En el experimento del dado, el conjunto de los números pares entre 1 y 6 es un evento. Si el dado arroja 2, 4 o 6, entonces ocurre este evento. Una ***medida de probabilidad*** consiste en asignar a cada elemento del espacio muestral un número entre 0 y 1 de modo que la suma de las probabilidades de todas las posibilidades es igual a 1. En el experimento "tirar una moneda", la posibilidad cara tiene probabilidad ½, al igual que la posibilidad sello. La suma de estas probabilidades es ½ + ½ = 1. La ***probabilidad de un evento*** es la suma de las probabilidades de cada uno de sus elementos. La probabilidad de un evento da una medida de cuán verosímil es la ocurrencia de dicho evento.

El espacio muestral, la colección de todos los eventos y la medida de probabilidad constituyen la base conceptual de la teoría de probabilidad. Juntos, constituyen el ***espacio de probabilidad***. En casos concretos, es muy importante identificar estas tres componentes.

Un poco de teoría de conjuntos: recordemos tres conceptos básicos de teoría de conjuntos en el lenguaje de las probabilidades. Dados dos eventos, su **unión** es el evento obtenido al juntar todos los elementos de ambos eventos. La **intersección** de dos eventos es el evento formado por todos los elementos que tienen en común. Si dos eventos no tienen elementos en común, decimos que son **disjuntos** y su intersección es el **conjunto vacío**. El **complemento** —o negación— de un evento es el evento formado por todos los elementos que no están en él.

Hay bastantes conceptos condensados en los párrafos anteriores. Aterricemos la discusión a una situación más familiar: el experimento "tirar un dado". Como ya señalamos, el espacio muestral consiste en los números enteros del 1 al 6. Un evento en este contexto es por lo tanto una colección de números del 1 al 6. Por ejemplo, el conjunto formado por los números 1, 2 y 5 es un evento. Que ocurra este evento significa que, al tirar el dado, el resultado es uno de estos números. La medida de probabilidad en este experimento le asigna probabilidad $\frac{1}{6}$ a cada una de las distintas posibilidades. La probabilidad del evento anterior —los números 1, 2 y 5— es la suma de las probabilidades de cada uno de estos números, es decir, $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 3/6 = \frac{1}{2}$. Otro ejemplo: la sentencia "el resultado del experimento es un número mayor o igual a 5" describe un evento. Este evento tiene dos posibilidades: el 5 y el 6. La probabilidad de este evento es igual a la probabilidad de obtener 5 más la probabilidad de obtener 6: $\frac{1}{6} + \frac{1}{6} = 2/6 = \frac{1}{3}$. En todos los ejemplos tratados en este libro, identificaremos cuidadosamente las tres componentes del espacio de probabilidad correspondiente.

El matemático ruso Andreï Kolmogorov (1903-1987) sentó las bases conceptuales de la teoría de probabilidad en 1933 al enunciar un conjunto de tres axiomas sobre los cuales se articuló la teoría de ahí en adelante. En general, las teorías matemáticas se formulan a partir de definiciones básicas que no son demostradas —los axiomas— sobre las cuales se cimientan las demostraciones de todos los teoremas, verdades inmutables que se deben demostrar rigurosamente. Los **tres axiomas de Kolmogorov** son:

1.- La probabilidad de un evento es un número positivo.

2.- La suma de las probabilidades de todos los elementos del espacio muestral es igual a 1.

3.- La probabilidad de la unión de eventos disjuntos —es decir, que no tienen elementos en común— es igual a la suma de sus probabilidades individuales.

La formulación anterior —y, en especial, el primer axioma— es extremadamente simple. El segundo y el tercer axioma no son tan inocentes, pues pueden referirse a infinitos elementos o eventos respectivamente. Sumar una cantidad infinita de números es un asunto delicado; pensemos en el griego Zenón y sus famosas paradojas. Con estos axiomas en mano, la probabilidad cuenta con la ayuda de una poderosa herramienta matemática: la teoría de la medida, una disciplina que ya había llegado a su madurez cuando Kolmogorov enunció sus axiomas.

> El matemático ruso **Andreï Kolmogorov (1903-1987)** es una de las personalidades más influyentes de la matemática del siglo XX. El alcance de sus contribuciones es prácticamente imposible de describir con la debida justicia y abarca la teoría de probabilidades, el análisis matemático, la lógica, la mecánica de fluidos, la mecánica clásica, los sistemas dinámicos, la computación y las matemáticas aplicadas. Numerosos artículos de Kolmogorov han sido recopilados en seis tomos de trabajos selectos y representan un tesoro de belleza matemática. Kolmogorov fue el mentor de algunos de los matemáticos rusos más relevantes y contribuyó ampliamente a la enseñanza de las matemáticas a través de sus libros.

La medida de probabilidad más sencilla le asigna la misma probabilidad a cada elemento del espacio de probabilidad. Esta se conoce como ***probabilidad uniforme***. En el caso del dado, cada posibilidad tiene probabilidad ⅙. Este es, por lo tanto, un ejemplo de probabilidad uniforme. El cara o sello y el día de cumpleaños son otros ejemplos de probabilidad uniforme, el primero sobre un espacio muestral muy pequeño —cara o sello— y el segundo sobre un espacio muestral bastante grande —todos los días del año—. Cuando una cantidad puede tomar un determinado número de valores, todos ellos con la misma probabilidad, también decimos que se distribuye uniformemente o que sigue la ***distribución uniforme***. También usamos el término *equiprobable*. *Probabilidad uniforme* y *distribución uniforme* son sinónimos, aunque nos inclinaremos por la segunda terminología en el capítulo sobre distribuciones de probabilidad.

En teoría de probabilidades, se suele expresar las probabilidades de manera simbólica. Así como la incógnita en una ecuación se denota habitualmente por la letra **x**, la probabilidad de un evento se denota genéricamente por la letra **p**. En el cara o sello, la probabilidad de cualquiera de las dos posibilidades en **p** = ½. Para un dado, la probabilidad de que salga el número 2 es **p** = ⅙.

Recordemos que dos eventos son disjuntos si no tienen elementos en común. En este caso, la probabilidad de la unión de estos eventos es simplemente la suma de sus probabilidades individuales. Un ejemplo genérico de esto es la relación entre un evento y su complemento. El complemento —o negación— de un evento es un evento. Por ejemplo, el complemento del evento "Lloverá mañana" es el evento "No lloverá mañana". En cualquier caso, o bien ocurre un evento o bien no ocurre, ambas posibilidades con cierta probabilidad. En otras palabras, o bien ocurre un evento o bien ocurre su complemento. Como la suma de las probabilidades tiene que ser igual a 1, la suma de la probabilidad de un evento y de la probabilidad de su complemento es igual a 1. Simbólicamente, si la probabilidad de un evento es igual a **p**, entonces, la probabilidad de su complemento es igual a 1 – **p**. En ocasiones nos referiremos al cálculo de la probabilidad del complemento de un evento como "paso al complemento".

Como ya sabemos, la probabilidad de un evento es la suma de las probabilidades de cada uno de sus elementos. Con la probabilidad uniforme, la probabilidad de un evento también puede calcularse como el cociente entre el **número de posibilidades favorables** y el **número de posibilidades totales**. Volviendo al experimento del dado, el evento "el resultado del experimento es un número mayor o igual a 5" consta de dos posibilidades favorables: el 5 y el 6. El número total de posibilidades es igual a 6. De este modo, la probabilidad de este evento es 2/6 = ⅓ ≈ 0,33, tal como concluimos anteriormente.

Por supuesto, hay situaciones muchísimo más complicadas. La sentencia "Lloverá mañana" describe un evento correspondiente a un experimento de difícil predicción: básicamente, dejar pasar el tiempo y observar si llueve o no llueve. El cálculo de la probabilidad de este evento involucra una gran cantidad de variables y no se limita a contar posibilidades favorables versus posibilidades totales.

Contamos ahora con una base sólida de herramientas teóricas para recorrer nuestro paseo por el azar.

Un ejemplo clásico: la urna

Veremos ahora un ejemplo clásico de probabilidad: la urna. La teoría de probabilidades consta de una serie de "problemas-tipos" que son sencillos de formular y abarcan un gran número de contextos. La urna, en sus dos variantes que estudiaremos, describe, por ejemplo, la lotería, las manos de póker o la paradoja del cumpleaños que estudiaremos en el próximo capítulo.

Describamos una situación concreta. Tenemos una urna con cinco bolas numeradas del 1 al 5, y sacamos las bolas una tras otra. Hay dos variantes de este problema: la urna con reposición y la urna sin reposición. En la urna con reposición, sacamos una bola y la volvemos a poner en la urna antes de sacar la siguiente. Por el contrario, en la urna sin reposición, no volvemos a poner la bola sacada en el primer sorteo antes de escoger la segunda y las bolas sorteadas son, por lo tanto, siempre distintas.

Como señalamos, la urna es un experimento bastante genérico. La lotería es una urna sin reposición con 41 bolas. En la paradoja del cumpleaños que veremos más adelante, la fecha de cumpleaños es una urna con 365 bolas que se considerará primero con reposición y luego sin reposición.

Antes de iniciar nuestra discusión de la urna, introduciremos un concepto fundamental de la teoría de probabilidades: la independencia. Decimos que dos eventos son **independientes** si la ocurrencia o no ocurrencia del primero no incide en la ocurrencia del segundo, y viceversa. Este es típicamente el caso cuando se realizan simultáneamente dos experimentos distintos. Por ejemplo, si lanzamos dos dados, el resultado del primer dado no incide en el resultado del segundo y cualquier evento relacionado solamente con el primer dado es independiente de cualquier evento relacionado solamente con el segundo. Cuando dos eventos son independientes, la probabilidad de que ocurran ambos es simplemente el producto de las probabilidades de cada uno de ellos. No se debe confundir eventos disjuntos y eventos independientes.

Recordemos que en la urna con reposición sacamos una bola y la volvemos a poner en la urna antes de sacar la siguiente. ¿Cuál es la probabilidad de sacar primero el número 2 y, luego, el número 4? En el primer sorteo, hay cinco posibilidades. La probabilidad de sacar el número 2 es ⅕. En el segundo sorteo, también hay cinco posibilidades. La probabilidad de sacar el número 4 es ⅕. Estos dos eventos son independientes. Sacar o no sacar el 2 en el primer sorteo no incide en el resultado del segundo sorteo. La probabilidad buscada es ⅕ × ⅕ = 1/25 = 0,04.

En la urna sin reposición, no volvemos a poner la bola sacada en el primero sorteo antes de escoger la segunda. ¿Cuál es la probabilidad de sacar primero el número 2 y, luego, el número 4? Nuevamente, hay cinco posibilidades en el primer sorteo y la probabilidad de escoger el número 2 es $\frac{1}{5}$. Sin embargo, dado que solo quedan cuatro bolas en la urna, hay solamente cuatro posibilidades para el segundo sorteo. La probabilidad de obtener el número 4 es $\frac{1}{4}$. Luego, la probabilidad de escoger el número 2 y luego el número 5 es igual a $\frac{1}{5} \times \frac{1}{4} = 1/20 = 0{,}05$.

En el caso de la urna con reposición, perdemos la independencia. El resultado del primer sorteo incide de alguna manera en el resultado del segundo, puesto que ya no puede volver a salir el mismo número.

Volvamos a calcular estas probabilidades como cociente de posibilidades favorables versus posibilidades totales. En el caso de la urna con reposición, nuestro espacio muestral consiste en todos los pares de números escogidos entre el 1 y el 5. Hay cinco posibilidades para el primero y cinco para el segundo. Hay entonces $5 \times 5 = 5^2 = 25$ posibilidades para escoger el par de números: el espacio muestral consta de 25 pares de números. De estas 25 posibilidades, nos interesaba solo una: el par 2 y 4 en ese orden. Su probabilidad es $1/25 = 0{,}04$ y recuperamos el número anterior.

En el caso de la urna sin reposición, hay cinco posibilidades para el primer número y cuatro posibilidades para el segundo. Hay $5 \times 4 = 20$ posibilidades totales que corresponden a pares ordenados de números del 1 al 5. La probabilidad de que salga cualquier par de números distintos en esta modalidad es $1 / 20 = 0{,}05$. Insistiremos en esta "multiplicatividad" del número de posibilidades en el capítulo 12, dedicado a la combinatoria.

La lotería es una urna sin reposición con 41 bolas numeradas del 1 al 41 de la cual se extraen seis bolas. Podemos intentar aplicar el razonamiento anterior para determinar el número de sorteos posibles. Hay, sin embargo, una sutileza crucial. Cuando contamos las posibilidades para la urna, consideramos secuencias ordenadas: primero el 2 y luego el 5. En un sorteo, poco importa el orden de aparición de las bolas. De cierta manera, nuestro método de conteo nos lleva a contar la misma configuración varias veces. Solucionaremos este problema en el capítulo 12 en nuestra discusión de las combinaciones y permutaciones.

La ley de los números grandes otra vez

Antes de seguir, reconciliaremos la noción de probabilidad como frecuencia con nuestro método de cálculo basado en el conteo de posibilidades.

Tomemos el ejemplo del dado. Para la ocurrencia de cualquiera de los resultados, hay seis posibilidades totales y solo una posibilidad favorable. Nuestro método para calcular probabilidades uniformes nos entrega el resultado ⅙: la probabilidad de que salga cualquier número dado en una tirada es igual a ⅙.

Notemos que hemos asumido de partida que la probabilidad en este caso es uniforme. Este es un supuesto razonable dada la simetría del dado. Ninguna cara parece ser privilegiada. Pero ¿cómo contrastar esto con la realidad? La clave está en la noción de frecuencia y la ley de los números grandes.

Si tiramos el dado una cantidad pequeña de veces, la combinación de números obtenida será del todo incoherente. Algunos números se repetirán, otros quizás no aparecerán. La ley de los números grandes nos dice que, al lanzar el dado una gran cantidad de veces, la proporción o frecuencia de aparición de cada número será cercana a ⅙. Tal como en el caso del cara o sello, puede ocurrir que algún número aparezca muchas más veces que otros. Lo importante es que esta diferencia es pequeña comparada con el número total de tiradas.

Esto es un principio general. Si un experimento puede resultar en **n** resultados distintos, cada uno con la misma probabilidad, entonces, al repetir el experimento un gran número de veces, la frecuencia de ocurrencia de cada una de estas posibilidades será muy cercana a 1/**n**. Aquí, **n** representa un número cuyo valor se debe deducir en cada caso. Para el dado, **n** = 6. Para el cara o sello, **n** = 2.

La ley de los números grandes nos permite poner a prueba la adecuada construcción de un dado. Si, al lanzarlo una gran cantidad de veces, la proporción de ocurrencia de un número es sistemáticamente mayor que ⅙, entonces, es muy probable que el dado tenga alguna imperfección que favorezca la aparición de ese número en particular. Decimos que el dado tiene un sesgo. En este caso, el dado no entrega una probabilidad uniforme sobre los números del 1 al 6. Discutiremos largamente acerca de dados sesgados en la próxima sección y acerca de la posibilidad o imposibilidad de sesgar una moneda.

¿Se pueden alterar las probabilidades de una moneda?

Todo partido de fútbol se inicia tras un lanzamiento de moneda que determina quién abrirá el juego. Este método de decisión es el paradigma de lo justo y objetivo. Quizás por esta razón causó tanto revuelo una noticia que anunciaba que la moneda belga de un euro resultaba en un cara o sello de probabilidades asimétricas.

La moneda sesgada —es decir, cuyas probabilidades de cara y sello no son iguales a ½— es el equivalente de un mito urbano en el mundo de la probabilidad. Numerosos ejemplos y problemas de textos guías clásicos empiezan con la frase "Lanzamos una moneda cuya probabilidad de cara es **p**, ...". Pero nadie nunca parece haber visto una moneda sesgada o verificado que una cierta moneda tenga algún sesgo, con una excepción notable: Kerrich, en 1946, construyó una "moneda" que aterrizó en cara 679 veces en 1.000 tiradas.

El problema de la moneda sesgada llevó a Gelman y Nolan a realizar en 2002 una serie de experimentos con estudiantes a quienes se invitaba a construir una moneda y un dado sesgados. Sesgar un dado es sencillo. La solución más eficaz fue encontrada por una estudiante que modificó la cara opuesta al 6 dándole una forma redondeada. Tras 120 lanzamientos del dado modificado, el 6 apareció una sola vez. Los experimentos con monedas fueron menos exitosos. La misma alumna construyó una moneda que aterrizó en sello 77 veces de 100 al lanzarla suavemente. Sin embargo, con lanzamientos más vigorosos, la cantidad de sellos bajó a 47, un resultado consistente con una probabilidad igual a ½.

El dado con una cara redondeada es un caso extremo. La forma más elemental de sesgar un dado es modificar su "centro de gravedad". El centro de gravedad de un cuerpo es el punto alrededor del cual la masa del objeto se distribuye equitativamente. En un dado perfecto, este se encuentra exactamente en el centro del cubo. Al agregarle masa a una de las caras, el centro de gravedad se desplaza hacia ella y la cara opuesta aumenta su probabilidad de ocurrir.

Este fue el procedimiento seguido por Kerrich. Construyó un disco de madera con una cara bañada en plomo, lo que desplaza efectivamente el centro de gravedad hacia la cara modificada. Otra vez, los lanzamientos de Kerrich fueron suaves.

La trayectoria de una moneda es gobernada por las ecuaciones de la mecánica de Newton. Estas ecuaciones describen de manera exacta el destino de un cuerpo en movimiento a partir de su posición inicial y de la velocidad que se le imprime; en el caso de la moneda, también de su velocidad de rotación. Estas cantidades se conocen como condiciones iniciales. Joseph Keller analizó el problema de la moneda en 1986 a partir de las ecuaciones de Newton y logró justificar de manera física la equiprobabilidad de caras y sellos.

Keller demostró que, en cierto rango, la mitad de las condiciones iniciales terminan en caras y la otra mitad en sellos. Además, mientras gira en el aire, la moneda pasa exactamente la mitad del tiempo cara arriba y cara abajo. La incerteza inicial dada por la imprecisión inherente del lanzado resulta en cara o sello, sin ninguna preferencia por el uno o el otro. Más aún, el eje de rotación de una moneda atraviesa su centro de gravedad, pero el movimiento resultante es independiente de su posición exacta. Esto explica que, en principio, una moneda modificada sigue teniendo probabilidades iguales a ½.

¿En qué radica entonces el resultado contradictorio del experimento de Kerrich? La clave está en el rango de condiciones iniciales en el que las conclusiones de Keller son válidas. El matemático y mago Persi Diaconis realizó mediciones en las que concluye que, típicamente, una moneda lanzada de la manera usual gira a una velocidad promedio de 38 rotaciones por segundo y realiza unas 19 revoluciones antes de tocar el suelo. Los resultados de Keller abarcan ampliamente este contexto. Kerrich, por otro lado, realizó lanzamientos con baja velocidad para los cuales las conclusiones de Keller no aplican. Lo mismo ocurre con los experimentos con la moneda de un euro. Se sospecha también que, a bajas velocidades de rotación, el rebote de una moneda modificada al golpear la superficie influye en el resultado del lanzamiento, favoreciendo una cara más que la otra.

Todo radica entonces en el modo de lanzar la moneda. Esto es consistente con los resultados arrojados por la moneda modificada de la estudiante de Gelman y Nolan. Lanzamientos suaves fuera del rango considerado por Keller dieron lugar a una probabilidad desigual de caras y sellos, mientras que lanzamientos vigorosos resultaron en la consabida probabilidad ½. Recordemos que un mago entrenado es capaz de sesgar el resultado de su lanzamiento, pero, otra vez, la velocidad de rotación imprimida a la moneda es relativamente baja.

Este paradigma aplica a casos extremos. Con mucha abstracción, podemos considerar que la tapa de un frasco es una moneda en un sentido amplio. Un matemático acostumbrado a las generalizaciones no tendrá ningún inconveniente con esta afirmación. Es, sin embargo, una moneda patológica, dado que su centro de gravedad se encuentra fuera de ella, en algún lugar del "hueco" interior de la tapa. Si lanzamos una tapa de modo que el borde golpea el suelo en primer lugar, la tapa se estabilizará sobre su cara plana con probabilidad muy cercana a uno. Sin embargo, un lanzamiento giratorio vigoroso nuevamente lleva a una probabilidad uniforme sobre ambas caras.

Esto nos lleva a una pregunta interesante. ¿Qué significa realmente que la probabilidad de cara sea igual a ½? En acuerdo con la discusión anterior, también debe explicitarse el modo de lanzamiento en el experimento. Esto revela una verdad más general: un experimento siempre debe ser diseñado de manera cuidadosa para lograr conclusiones correctas y acotar el rango de validez de estas conclusiones.

¿Cómo se debe lanzar una moneda? Simplemente de la manera en la que todos lo hacemos. Más aún, dejarla caer al suelo o atraparla en el aire y voltearla sobre el antebrazo no conduce a ninguna diferencia. Que aquellos preocupados por el inicio de los partidos de fútbol se queden tranquilos.

Todo lo probable ocurrirá alguna vez

Uno de los problemas clásicos de probabilidades se conoce como el *problema del simio*. Se trata de un experimento mental que se ilustra imaginando un simio que oprime aleatoriamente las teclas de una máquina de escribir, un artefacto ciertamente anacrónico que revela la antigüedad del problema. La producción literaria del simio es mayormente incoherente, pero, si escribe eternamente, en algún momento escribirá la totalidad del *Quijote*. Es más, la escribirá infinitas veces y con infinitas variantes ni siquiera imaginadas por Cervantes.

Este es un paradigma de la probabilidad: todo aquello que tiene probabilidad positiva de ocurrir ocurrirá alguna vez en un experimento repetido infinitas veces de manera independiente. La trampa, por supuesto, es que ningún experimento se repite infinitas veces y, en el caso del simio escritor, aunque este procedimiento se llevara a cabo de un momento en adelante sin interrupción, el tiempo necesario para observar las primeras líneas del *Quijote* podría ser mayor a la edad del universo. El problema del simio consiste en estimar este tiempo de espera aleatorio.

Este problema nos recuerda la célebre *Biblioteca de Babel* de Borges. En palabras de Borges, el universo —que otros llaman La Biblioteca— consta de una concatenación de habitaciones hexagonales que contienen libros de 410 páginas de 40 renglones de unas 80 letras, formados por todas las posibles combinaciones de 25 caracteres: 22 letras, el espacio, el punto y la coma. Borges relata que cantidades de hombres recorrieron centenares de habitaciones en busca de un libro coherente sin encontrarlo jamás. A lo más se encontró alguna frase; "*Oh tiempo tus pirámides*" es la única que se menciona. Como en el problema del simio, la mayoría de las combinaciones posibles de letras no tiene ningún sentido. Sin embargo, en algún lugar debe hallarse el *Quijote* y, citando el cuento, el catálogo de la Biblioteca, miles de catálogos falsos e incluso la autobiografía de los arcángeles.

Borges no clarifica si se trata de un universo finito con todas las combinaciones de libros o de un universo infinito con libros aleatorios. En ambos casos, todos los libros posibles se encuentran en algún lado.

Borges, en su fascinación por el infinito, postula la idea de un libro que contiene todos los libros. En ese caso, ¿no será el azar, caricaturizado por un simio frente a una máquina de escribir, este libro mitológico?

Probabilidades no uniformes y probabilidades continuas

Como vimos en el caso del dado sesgado y como veremos más adelante con mayor detalle, no todas las medidas de probabilidad son uniformes. Consideremos otro ejemplo: si nuestro experimento consiste en tirar un dado y sumar los números, el espacio muestral consiste en los números entre el 2 y el 12. Ahora, las probabilidades de ocurrencia de estos números son distintas. Para lograr una suma igual a 12, hay solo una posibilidad: que ambos dados muestren el número 6. Sin embargo, para formar el número 7, hay más posibilidades: $7 = 1 + 6 = 2 + 5 = 3 + 4 = 4 + 3 = 5 + 2 = 6 + 1$. Aquí, el primer y segundo término de la suma representan el resultado del primer dado del segundo respectivamente. En el caso del 7, hay seis posibilidades favorables. La probabilidad de que la suma sea igual a 7 es entonces seis veces mayor que la probabilidad de que sea igual a 12.

Si bien este ejemplo ilustra una medida de probabilidad no uniforme, se construyó a partir de probabilidades uniformes. Su estudio se reduce, al fin y al cabo, a contar posibilidades favorables y sumar probabilidades.

Existen casos en los que las probabilidades no se pueden calcular contando casos. Pensemos en el lanzamiento de jabalina. La distancia a la que un deportista lanza una jabalina es una cantidad que se extiende continuamente dentro de cierto rango. La probabilidad de que la jabalina llegue exactamente a una distancia de 75 metros es igual a 0. Y así con cualquier número fijo. Las posibles distancias alcanzadas por la jabalina son infinitas. No es posible obtener una medida de probabilidad contando casos favorables y casos posibles.

Lo que es posible, sin embargo, es calcular la frecuencia con la que la jabalina cayó entre los 75 y los 76 metros y, en general, en cualquier intervalo razonable. Esto nos permite establecer una medida de probabilidad aproximada para la distancia.

Veremos más adelante cómo representar las probabilidades no uniformes mediante gráficos y cómo estos gráficos nos dan un indicio de lo que son las probabilidades continuas.

Apéndice: más propiedades de las medidas de probabilidad

Discutiremos a continuación algunas propiedades elementales de las medidas de probabilidad que se derivan de los axiomas de Kolmogorov y de la noción de independencia. Nos interesaremos particularmente en las operaciones que se pueden realizar con eventos. Este material se incluye para profundizar nuestra discusión sobre los elementos básicos de la teoría de probabilidad y no es imprescindible para la lectura del resto del libro. El lector poco interesado en estas divagaciones matemáticas puede dirigirse al capítulo siguiente.

Recordemos que el complemento —o negación— de un evento es un evento. En el juego de cara o sello, la probabilidad del evento "sacar cara" es $\mathbf{p} = ½$. La probabilidad de la negación de este evento, es decir, "no sacar cara", es $1 - \mathbf{p} = 1 – ½ = ½$. Esto es coherente dado que el evento "no sacar cara" es exactamente el evento "sacar sello". En el dado, la probabilidad de sacar el número 3 es $\mathbf{p} = ⅙$. La probabilidad de no sacar el número 3 es $1 – \mathbf{p} = 1 - ⅙ = ⅚ \approx 0{,}83$.

Los eventos suelen denotarse por letras mayúsculas. Si A es un evento, su complemento se denota por A^c y corresponde al evento "No ocurrió A". Considerar el complemento de un evento se denomina coloquialmente como "pasar al complemento". Si A y B son dos eventos de un mismo espacio muestral, el evento "ocurre A o B" consta de todos los elementos de A y de todos los elementos de B. En término matemáticos, este evento es la **unión** de los eventos A y B. Por ejemplo, en el dado, si A es el evento "el número es par" y B es el evento "el número es mayor o igual a 5", el evento "ocurre A o B" es el evento "el número es par o el número es mayor o igual a 5". El evento A consta de los números 2, 4 y 6, y B de los números 5 y 6. Por lo tanto, el evento "ocurre A o B" consta de los números 2, 4, 5 y 6, y su probabilidad es 4/6 = ⅔ ≈ 0,66. Notamos que el número 6 pertenece a A y a B. Cuando formamos el evento "ocurre A o B", lo contamos una sola vez. El evento "ocurre A o B" se enuncia a menudo como "A o B" para mayor simplicidad.

Como señalamos anteriormente, cuando dos eventos no tienen elementos en común, decimos que son ***disjuntos***. En este caso, la probabilidad de su unión es la suma de las probabilidades de cada evento. Volviendo al dado, digamos que el evento C está formado por los números 1 y 3, y el evento D, por los números 2, 4 y 6. El evento "C o D" está formado por los números 1, 2, 3, 4 y 6. La probabilidad de C es 2/6. La probabilidad de D es 3/6. Finalmente, la probabilidad del evento "ocurre C o D" es ⅚ ≈ 0,83, que coincide con 2/6 + 3/6, que es la suma de la probabilidad de C y de la probabilidad de D.

Si E y F son dos eventos, el evento "ocurre E y F" o, simplemente, "E y F" consta de todos los elementos comunes a E y F. Es el evento "ocurren E y F simultáneamente". En términos matemáticos, corresponde a la ***intersección*** de los eventos E y F. Volviendo al dado, digamos que E es el evento "el número es menor o igual a 3" y F es el evento "el número es impar". El primer evento consta de los números 1, 2 y 3. El segundo, de los números 1, 3 y 5. La intersección, es decir, las posibilidades o elementos comunes a ambos eventos, está formada por los números 1 y 3. La probabilidad del evento "E y F" es 2/6 = ⅓ ≈ 0,33.

Recordemos que no deben confundirse eventos disjuntos y eventos independientes. En el caso de eventos independientes, la probabilidad de su intersección es el producto de sus probabilidades. Por ejemplo, supongamos que tiramos una moneda y luego un dado. Llamamos G al evento "obtenemos cara al lanzar la moneda" y H al evento "obtenemos el número 5 al lanzar el dado". El evento "G y H" es el evento "obtenemos cara con la moneda y 5 con el dado". Como G y H resultan de experimentos que no se relacionan de ningún modo, son eventos independientes. La probabilidad de "G y H" es el producto de las probabilidades de G y de H: ½ × ⅙ = 1/12 ≈ 0,08. Recordemos los eventos E y F del párrafo anterior para un solo dado: E es el evento "el número es menor o igual a 3" y F es el evento "el número es impar". E consta de dos números impares y de un número par. Luego, si ocurre E, es bastante probable que también ocurra F. Estos eventos no son independientes. Concretamente, la probabilidad de E es 3/6, la probabilidad de F es 3/6, y la probabilidad de la intersección de E y F es simplemente la probabilidad de obtener 1 o 3, que es igual a $2/6 \approx 0{,}33$ y que no corresponde al producto $3/6 \times 3/6 = 0{,}25$. Notemos que la intersección de dos eventos disjuntos es igual al conjunto vacío y tiene probabilidad 0.

Eventos que no son disjuntos: si dos eventos no son disjuntos, no resulta tan fácil aparentemente calcular la probabilidad de su unión. Un método sencillo consiste en calcular el número de posibilidades favorables para A, sumarle el número de posibilidades favorables para B y restar el número de posibilidades que fueron contadas dos veces, es decir, el evento "A y B".

CÁLCULO DE PROBABILIDADES

Si escogemos dos personas al azar, la probabilidad de que cumplan años el mismo día es bajísima. Lo mismo con tres personas. Pero ¿cuántas personas se necesita congregar para que la probabilidad de que dos de ellas compartan el mismo cumpleaños sea alta?

La respuesta a esta pregunta es sorprendente, tal como ocurre con las numerosas paradojas de las probabilidades, situaciones sencillas que desafían la intuición. Ya hemos mencionado que nuestra intuición probabilística es bastante pobre. De ahora en adelante, disponemos de herramientas de cálculo para llegar a conclusiones rigurosas que, probablemente, no dejarán de sorprendernos.

El juego de dados del Chevalier de Méré

Calcularemos las probabilidades involucradas en los juegos de un dado y de dos dados del Chevalier de Méré.

El primer juego es el siguiente: los jugadores A y B apuestan la misma cantidad de dinero. Se tira un dado cuatro veces. Si sale al menos un 6, gana el jugador A. En caso contrario, gana el jugador B.

Resulta que las probabilidades de ganar están levemente a favor del jugador A. De Méré llegó a esta conclusión siguiendo un razonamiento incorrecto. Sin embargo, la conclusión es correcta, como probablemente De Méré lo habrá corroborado jugando este juego repetidamente; le bastó comprobar que ganaba más de un 50% de las veces y utilizar su intuición, aun si no conocía la ley de los números grandes.

De Méré razonó de la manera siguiente: sacar un 6 en una jugada tiene probabilidad $\frac{1}{6} \approx 0{,}17$; luego, sacar un 6 en cuatro jugadas es $4 \times \frac{1}{6} = \frac{2}{3} \approx 0{,}66$. Si bien esto es efectivamente mayor que $\frac{1}{2}$, no es la respuesta correcta. No podemos culpar a De Méré. No disponía de los métodos necesarios para realizar el cálculo.

Calculemos la probabilidad real con todo detalle. Curiosamente, ¡resulta más fácil calcular la probabilidad de que el jugador A pierda! Esto es bastante común en probabilidades. A veces, es más fácil calcular la probabilidad de que un evento no ocurra, o sea, la probabilidad de su **complemento**. Una vez que tenemos la probabilidad del complemento, es sencillo calcular la probabilidad del evento que nos interesa: es igual a 1 menos la probabilidad del complemento. Por ejemplo, si el complemento de un evento tiene probabilidad 25%, entonces, el evento mismo tiene probabilidad 100% – 25% = 1 – 0,25 = 0,75 = 75%.

La probabilidad de no sacar 6 en la primera tirada es $\frac{5}{6}$. Hay cinco posibilidades "favorables", que son los números del 1 al 5, contra seis posibilidades totales. La probabilidad de no sacar 6 en la segunda jugada es otra vez $\frac{5}{6}$, y así con cada jugada. Todas estas jugadas son independientes: el sacar un 6 o no en la primera jugada no incide en nuestra probabilidad de sacar un 6 en las siguientes. Luego, la probabilidad de no sacar 6 en ninguna de las cuatro tiradas es $\frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = 625/1.296 \approx 0{,}48$. La probabilidad de sacar al menos un 6 en cuatro tiradas es la probabilidad del complemento de este evento, o sea, aproximadamente 0,52. Concluimos que la probabilidad de que gane A es levemente mayor a $\frac{1}{2} = 0{,}50$.

El número total de jugadas igual a 4 no es casual. 4 es el número más pequeño tal que la probabilidad de que gane A es mayor que $\frac{1}{2}$. Con una sola tirada, la probabilidad de que gane A es $\frac{1}{6}$. El razonamiento del párrafo anterior nos permite calcular la probabilidad correspondiente en el caso de dos y tres jugadas.

El segundo juego propuesto por el Chevalier es muy similar, pero involucra tiradas repetidas de dos dados. Nuevamente, los jugadores A y B apuestan la misma cantidad de dinero, pero, esta vez, A gana si sale un doble 6 en una cierta cantidad de jugadas. El Chevalier estaba interesado en determinar cuántas tiradas se debían efectuar para que la probabilidad de ganar sea mayor a $\frac{1}{2}$. De Meré pensaba que esta cantidad es igual a 24.

El razonamiento que llevó a De Méré a esta conclusión es incorrecto. Curiosamente, aplicando este mismo razonamiento al caso de dos dados, obtuvo una respuesta parcialmente correcta, es decir, mayor que $\frac{1}{2}$.

Recordemos el razonamiento del Chevalier: en el caso de un solo dado, la probabilidad de sacar un 6 es $\frac{1}{6} \approx 0{,}17$. Luego, según De Méré, la probabilidad de sacar un 6 en cuatro jugadas es 4 × ⅙ = ⅔ ≈ 0,66, que es mayor que ½ = 0,50. Más arriba, calculamos esta probabilidad y concluimos que no es igual a ⅔ ≈ 0,66. Por lo tanto, el argumento del Chevalier es incorrecto.

Apliquemos, sin embargo, este mismo razonamiento al juego con dos dados. La probabilidad de sacar un doble 6 es ⅙ × ⅙ = 1/36 ≈ 0,03. Con 24 tiradas, la probabilidad de sacar algún doble 6 debe ser, según De Méré, igual a 24 × 1/36 ≈ 0,66. Este resultado no solo es incorrecto, sino que la probabilidad correcta es menor que ½. En este caso, las conclusiones de De Méré deben haber tenido consecuencias trágicas, ya que, al jugar este juego, se arriesgaba a perder sistemáticamente. Por la ley de los números grandes, sabemos que la frecuencia de juegos ganados debe acercarse a la probabilidad real de ganar. Así, al observar que perdía más del 50% de las veces, De Méré intuyó que su cálculo era incorrecto y contactó al matemático Blaise Pascal. Repetimos: si bien no conocía la ley de los números grandes, pudo darse cuenta de que algo andaba mal.

La manera correcta de calcular la probabilidad de ganar es idéntica al caso de un solo dado. La probabilidad de sacar un doble 6 en una tirada de dos dados es 1/36 ≈ 0,03. Luego, la probabilidad de no sacar un doble 6 en una tirada es 1 – 1/36 = 35/36 ≈ 0,97. Luego, la probabilidad de no sacar un doble 6 en 24 tiradas es 35/36 × 35/36 × …… × 35/36, donde el producto consiste en 35/36 repeticiones del factor 35/36, es decir $(35/36)^{24}$. El resultado de este producto es aproximadamente 0,51. Luego, la probabilidad de sacar un doble 6 en 24 tiradas es aproximadamente 1 − 0,51 = 0,49. Vemos que esta probabilidad es menor que ½ = 0,50. En este juego, De Méré ganaría aproximadamente 49 veces de cada 100.

Con un cálculo similar, se puede concluir que es necesario arrojar el dado 25 veces para tener una probabilidad mayor que ½ de ganar. Con 25 tiradas, la probabilidad de no sacar ningún doble 6 es aproximadamente igual a 0,49, por lo tanto, la probabilidad de ganar es aproximadamente 1 – 0,49 = 0,51, que es —un poco— mayor que ½ = 0,50. El lector podrá verificar este cálculo con el método del párrafo anterior.

La paradoja del cumpleaños, versión simplificada

La probabilidad lleva a situaciones sorprendentes. La paradoja del cumpleaños es un ejemplo clásico. La pregunta es la siguiente:

¿cuántas personas se necesitan para que la probabilidad de que dos de ellas tengan cumpleaños el mismo día sea mayor que ½? Responderemos esta pregunta en la próxima sección.

Simplificaremos la situación por ahora: ¿cuántas personas se necesitan para que la probabilidad de que dos de ellas tengan cumpleaños un mismo día de la semana sea mayor que ½?

Vamos a suponer que la probabilidad de que alguien nazca un cierto día de la semana es ⅐. Así, la probabilidad de nacer un lunes es igual a ⅐, al igual que la probabilidad de nacer un jueves o cualquier otro día.

¿Cuál es la probabilidad de que dos personas nazcan el mismo día de la semana? Como en el juego de dados de De Méré, es más fácil calcular la probabilidad del complemento: ¿cuál es la probabilidad de que dos personas no nazcan el mismo día de la semana? Para el cumpleaños de una persona, tenemos siete posibilidades. Para el cumpleaños de dos personas, tenemos por lo tanto $7 \times 7 = 49$ posibilidades. Algunas de estas posibilidades brindan cumpleaños el mismo día. Las tenemos que descartar. Veamos de cuántas maneras distintas podemos ubicar los cumpleaños de dos personas de modo que no coincidan. Tenemos siete posibilidades para la primera persona. Una vez que fijamos este cumpleaños, quedan solo seis posibilidades para la segunda persona. Luego, hay $7 \times 6 = 42$ posibilidades para que dos personas cumplan años un día distinto de la semana. La probabilidad de que dos personas cumplan años un día distinto de la semana es, por lo tanto, igual a $42/49 \approx 0{,}86$. Volviendo a la pregunta original, la probabilidad de que dos personas cumplan años el mismo día de la semana es aproximadamente igual a $1 - 0{,}86 = 0{,}14$.

Para tres personas, el razonamiento es el mismo. Hay $7 \times 7 \times 7 = 343$ posibilidades para acomodar el cumpleaños de tres personas, pero solo $7 \times 6 \times 5 = 210$ corresponden a cumpleaños en días distintos. Esto nos da una probabilidad de $210/343 \approx 0{,}61$ para que tres personas tengan cumpleaños en días diferentes de la semana. Luego, la probabilidad de que dos de tres personas cumplan años un mismo día de la semana es aproximadamente igual a $1 - 0{,}61 = 0{,}39$. Esto aún es menor que ½.

El cálculo para cuatro personas es similar. La probabilidad de que cuatro personas cumplan año en días distintos de la semana es igual a $840/2.401 \approx 0{,}35$, donde $840 = 7 \times 6 \times 5 \times 4$ y $2.401 = 7 \times 7 \times 7 \times 7$. La probabilidad de que al menos dos de cuatro personas cumplan años el mismo día de la semana es aproximadamente igual a $1 - 0{,}35 = 0{,}65$, que es un número mayor que ½. La respuesta a nuestra pregunta es 4.

Notemos que la probabilidad de que, en un grupo de ocho personas, dos de ellas cumplan años un mismo día de la semana es igual a 1. Esto se debe al **principio del palomar**. Si tenemos que repartir ocho personas en siete categorías, necesariamente dos de ellas deben encontrarse en la misma. El lector podrá verificar que la probabilidad de que, en un grupo de siete personas, dos de ellas cumplan años un mismo día de la semana es bastante cercana a 1.

La paradoja del cumpleaños

Recordemos la pregunta: ¿cuántas personas se necesitan para que la probabilidad de que dos de ellas tengan cumpleaños el mismo día sea mayor que ½? En la sección anterior, remplazamos los días del año por los días de la semana para simplificar el problema. Ahora, contestaremos la pregunta original.

El día de cumpleaños proporciona otro ejemplo de probabilidad uniforme. Para simplificar la situación, no tomaremos en consideración los años bisiestos y supondremos que un año tiene 365 días. Asumiremos el supuesto razonable de que la probabilidad de nacer en cualquier día dado —por ejemplo, el 23 de febrero— es uniformemente igual a 1/365. Este es un número bastante pequeño: es aproximadamente igual a 0,003. Esto es apenas tres milésimas, es decir, 0,3%. En otras palabras, de cada mil personas, solo tres de ellas en promedio nacieron el 23 de febrero.

¿Cuál es la probabilidad de que dos personas tengan el mismo cumpleaños? Como en el problema de los dados del Chevalier de Méré y en el ejemplo anterior, es más fácil calcular la probabilidad de que no tengan el mismo cumpleaños. Hay 365 posibilidades para un cumpleaños. Por lo tanto, hay $365 \times 365 = 365^2 = 133.225$ posibilidades para dos cumpleaños. Esto es exactamente lo que ocurre en una urna con 365 bolas con reposición. Contemos ahora las posibilidades para que dos personas no tengan el mismo cumpleaños. Para el cumpleaños de la primera persona, tenemos 365 posibilidades. Una vez que seleccionamos una de ellas, quedan 364 posibilidades para que la segunda persona tenga un cumpleaños distinto de la primera. Luego, dado el carácter multiplicativo del conteo de posibilidades, hay $365 \times 364 = 132.860$ posibilidades para que dos personas tengan cumpleaños distintos. Luego, la probabilidad de que dos personas tengan distinto cumpleaños es $132.860/133.225 \approx 0,997$. Este es un número muy cercano a 1. Por esto, tuvimos que agregar un decimal en el resultado para no redondear directamente a 1. En porcentaje, esto es 99,7%. En otras palabras, es sumamente probable que dos personas cualesquiera cumplan año en días distintos. Pasando al complemento, la probabilidad de que dos personas tengan el mismo cumpleaños es aproximadamente 1 – 0,997 = 0,003 = 0,3%. Esto es un número muy pequeño.

¿Cuántas personas se necesitan para que la probabilidad de que al menos dos de ellas cumplan años el mismo día sea mayor que ½? La respuesta es un número sorprendentemente pequeño: 23.

Para calcular esta probabilidad, solo tenemos que seguir con el razonamiento del párrafo anterior. Hay 365 posibilidades para el cumpleaños de la primera persona. Una vez que fijamos esta posibilidad, quedan 364 posibilidades para el cumpleaños de la segunda persona. Para el cumpleaños de la tercera persona, tenemos que descartar las dos posibilidades ocupadas por las dos primeras personas: quedan 363 posibilidades. Para la cuarta persona, quedan 362 posibilidades. Procediendo de esta forma, vemos que, para la vigesimotercera persona, quedan 343 posibilidades. Multiplicando estas posibilidades, vemos que el número de posibilidades para que 23 personas cumplan años en días distintos es igual a $365 \times 364 \times 363 \times \ldots \times 345 \times 344 \times 343$. Este es un número con 59 dígitos que ni siquiera escribiremos. Sin embargo, el número de posibilidades totales es increíblemente grande, de modo que el cociente de estos dos números es un número entre 0 y 1. Concretamente, el número de posibilidades para los cumpleaños de 23 personas es 365 multiplicado consigo mismo 23 veces: 365^{23}. Luego, la probabilidad de que 23 personas cumplan años en días distintos es igual a:

$$365\times364\times363\times\ldots\times345\times344\times343/365^{23}\approx0,49$$

Volviendo a nuestro problema inicial, vemos que la probabilidad de que, en un grupo de 23 personas, al menos dos de ellas tengan su cumpleaños en un mismo día es aproximadamente igual a 1 – 0,49 = 0,51. ¡Este número es mayor que ½!

DISTRIBUCIONES DE PROBABILIDAD

¿Cuál es la probabilidad de obtener un 7 al sumar el resultado arrojado por dos dados? ¿Cuál es la probabilidad de obtener un *full house* en el póker? ¿Cuántos ataques de tiburones se producen cada un año? ¿Cuál es la posición de una partícula de polen suspendida en un líquido? ¿Cómo detectar fraudes en declaraciones de impuestos?

Todas estas preguntas nos llevan más allá de la distribución uniforme y a discutir las distribuciones de probabilidad con mayor generalidad.

En este capítulo, estudiaremos las medidas de probabilidad más importantes que describen, por ejemplo, las fluctuaciones de los precios de las acciones en la bolsa o la distribución de los dígitos más frecuentes en datos de la vida cotidiana.

Suma de dados

Un antiguo juego de dados consistía en sumar el resultado de tres dados. Era de conocimiento común que existen seis maneras distintas de obtener una suma igual a 9 y seis maneras distintas de obtener una suma igual a 10:

$$9=1+2+6=1+3+5=1+4+4=2+2+5=2+3+4=3+3+3$$
$$10=1+3+6=1+4+5=2+2+6=2+3+5=2+4+4=3+3+4$$

Basado en este cálculo, es muy razonable esperar que la probabilidad de obtener un 9 sea idéntica a la probabilidad de obtener un 10. Un jugador aficionado a este juego podrá fácilmente intuir que este no es el caso. Tras una gran cantidad de tiradas de tres dados, el jugador notará sin duda que el número 10 aparece más frecuentemente que el 9. ¿En qué radica esta aparente paradoja? El problema llegó a oídos de Galileo, quien rápidamente dio con la explicación.

Antes de resolver este enigma, estudiaremos lo que ocurre con la suma de dos dados. Esto nos permitirá iniciar nuestra discusión de las **distribuciones de probabilidades**, una de las cuales discutimos con bastante detalle en el capítulo anterior: la probabilidad uniforme. La suma de dos dados puede tomar cualquier valor entre 2 y 12. Las probabilidades de ocurrencia de cada uno de estos números constituyen la distribución de probabilidad asociada a la suma de dos dados. Ya veremos que no coincide con la probabilidad uniforme. Se utilizan los términos *distribución de probabilidad* y *ley de probabilidad* de manera intercambiable.

La siguiente tabla consigna los valores de la suma de dos dados según el resultado de cada uno de los dados. Los valores del primer dado se encuentran en la primera columna y los valores del segundo, en la primera fila.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

El espacio muestral de este experimento consiste en todos los números enteros entre el 2 y el 12. Sin embargo, vemos que el conjunto de todas las posibilidades para el resultado de dos dados es un espacio muestral mucho mayor que consta de todos los pares ordenados de números del 1 al 6. Todos estos pares tienen la misma probabilidad de ocurrir, por lo tanto, al lanzar dos dados, obtenemos la distribución uniforme sobre estos pares. Como hay $6 \times 6 = 36$ pares posibles, cada uno tiene una probabilidad de $1/36 \approx 0{,}03$ de ocurrir. La distribución de la suma, por otro lado, dista de ser uniforme.

Por ejemplo, solo hay una manera de obtener una suma igual a 2 o a 12. Ambos dados tienen que arrojar un 1 o un 6 respectivamente. Por lo tanto, las probabilidades de 2 y de 12 son ambas iguales a 1/36. En el caso extremo, vemos que hay seis maneras distintas de obtener un 7: $1 + 6 = 2 + 5 = 3 + 4 = 4 + 3 = 5 + 2 = 6 + 1$. Calculando el cociente entre los 6 casos favorables y los 36 casos posibles, calculamos que la probabilidad de obtener un 7 es igual a $6/36 = 1/6 \approx 0{,}17$. Como las probabilidades de obtener 2 y de obtener 7 son distintas, la distribución de la suma de dados no es uniforme. El lector podrá calcular fácilmente la probabilidad de ocurrencia de los resultados restantes basado en la tabla de más arriba.

Acercándonos a la paradoja de los tres dados, podríamos notar que hay tres maneras esencialmente distintas de obtener un 6 o un 7:

$$6=1+5=2+4=3+3$$
$$7=1+6=2+5=3+4$$

Sin embargo, una inspección de la tabla revela que la probabilidad de 6 es de solo 5/36 y, por lo tanto, menor que la probabilidad de 7. Esta aparente contradicción se resuelve observando que todos los pares que suman 7 tienen dos maneras distintas de ocurrir si tomamos en cuenta cuál de los dados arrojó los términos de la suma. La suma $7 = 1 + 6$ se puede lograr con el primer dado en 1 y el segundo en 6, y viceversa. Lo mismo ocurre con las otras posibles sumas que resultan en 7: todas descienden de dos posibilidades reales que aparecen en la tabla como dos posibilidades distintas. Lo mismo ocurre también con las sumas $1 + 5$ y $2 + 4$, ambas iguales a 6. La suma $3 + 3$, por otro lado, solo puede ocurrir de una manera: ambos dados tienen que resultar en 3. Esta es una única posibilidad en la tabla y explica por qué el 7 tiene mayor probabilidad de ocurrir que el 6.

El misterio de los tres dados fue resuelto de manera análoga por Galileo. Tres de las sumas que resultan en 9 o en 10 están dadas por tres números distintos. Cada una de estas combinaciones puede presentarse de seis maneras distintas. Por ejemplo, la suma $1 + 2 + 6$ se puede lograr como (1, 2, 6), (1, 6, 2), (2, 1, 6), (2, 6, 1), (6, 1, 2) y (6, 1, 2), donde hemos ordenado los números según aparecieron en los dados número 1, 2 o 3. Tres de las sumas que resultan en 10 y dos de las sumas que resultan en 9 tienen un número repetido. Como en el caso anterior, se puede verificar que cada una tiene tres maneras distintas de ocurrir si se considera el orden de los dados. La diferencia entre el 9 y el 10 radica en la suma $9 = 3 + 3 + 3$, que tiene solo una manera de ocurrir. Por lo tanto, la probabilidad de 9 es menor que la probabilidad de 10. Simplemente, el 9 tiene menos maneras distintas de ocurrir.

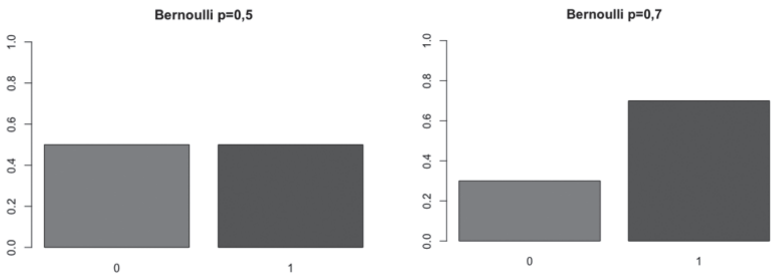
Distribuciones y sus gráficos

Dada cualquier colección de objetos, le podemos asignar una probabilidad a cada uno de ellos con la única restricción de que estas cantidades tienen que ser positivas y sumar uno. Esto define una **distribución de probabilidad** que, en principio, no tiene por qué ser uniforme. Por ahora, hemos considerado espacios muestrales finitos. Las distribuciones de probabilidad sobre estos espacios se conocen como distribuciones de probabilidad **discretas**. Recordemos que los términos *ley* y *distribución de probabilidad* son sinónimos.

La primera distribución de probabilidad que estudiamos proviene del juego de cara o sello. Ambas posibilidades, cara y sello, ocurren con probabilidad ½. El color de la ruleta sigue el mismo principio: rojo o negro. Hay otras maneras de asignarle una distribución de probabilidad a un espacio muestral de dos objetos. Si le asignamos una probabilidad **p** al primero, solo podemos asignarle una probabilidad 1 – **p** al segundo, lo que lleva a una distribución que no es uniforme. Estas distribuciones de probabilidad se conocen como **distribuciones de Bernoulli** en honor al matemático suizo Jacques Bernoulli (1654-1705). Una distribución de Bernoulli queda completamente caracterizada al dar el valor de **p**.

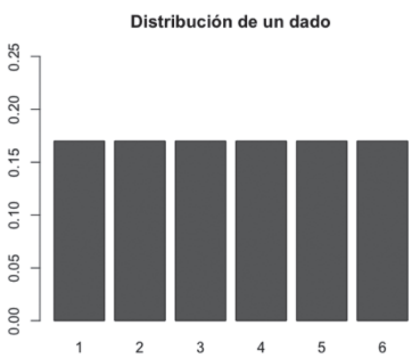
> **La familia Bernoulli**: algunas disciplinas están marcadas por las contribuciones de familias enteras. Es el caso de la familia Bach en la música. En matemáticas, este lugar de honor es ocupado por la familia suiza Bernoulli. Sus exponentes desarrollaron ideas fundamentales en la matemática y la física de los siglos XVII y XVIII. Johan Bernoulli fue también el mentor de uno de los matemáticos más relevantes de la historia: Leonhard Euler.

Las distribuciones de probabilidad se pueden representar mediante un gráfico. En el **gráfico de una distribución**, se ordenan los distintos elementos del espacio muestral sobre un segmento horizontal. Sobre cada una de estas ubicaciones, se dibuja una barra de área proporcional a la probabilidad de ocurrencia del elemento correspondiente. Los siguientes gráficos representan distribuciones de Bernoulli para distintos valores de **p**. Para simplificar, se consideró el espacio muestral consistente en dos elementos: **0** y **1**.

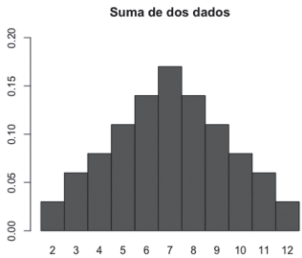


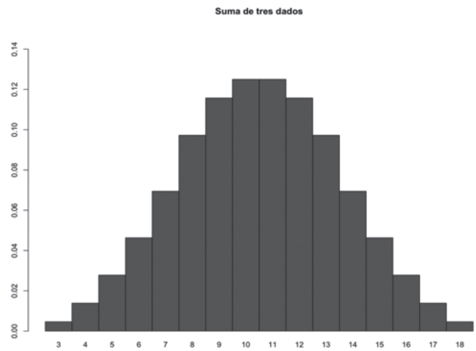
En una distribución de Bernoulli de parámetro ½ —como en cara o sello— ambas barras tienen la misma altura. Si el parámetro de la distribución es distinto a ½, obtenemos barras de tamaño distinto.

Una distribución uniforme tiene un gráfico completamente plano, dado que cada elemento del espacio muestral tiene la misma probabilidad de ocurrir, lo que se traduce en barras idénticas. El siguiente gráfico ilustra las probabilidades resultantes de arrojar un dado. Todas las barras tienen una altura igual a ⅙ ≈ 0,17.



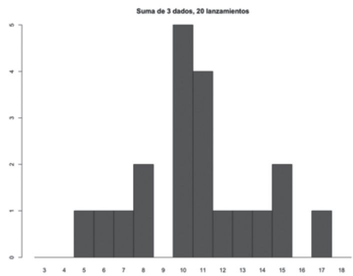
Los siguientes gráficos representan las distribuciones de probabilidad para la suma de dos y tres dados respectivamente. Como vemos, difieren de la distribución uniforme sobre los espacios muestrales correspondientes. Esto se infiere del distinto tamaño de las barras.

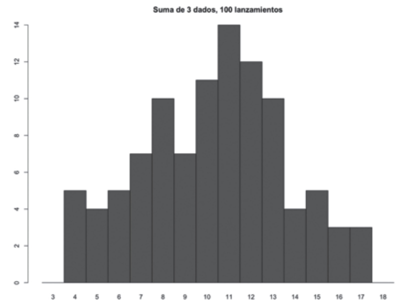


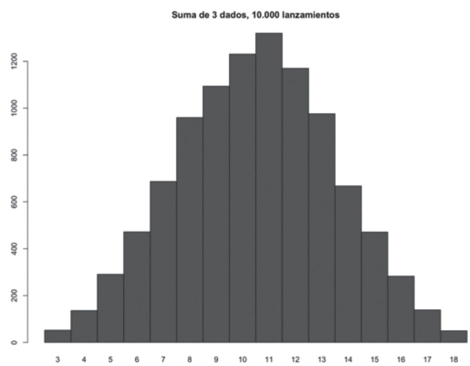


En la suma de dos dados, los números 2 y 12 tienen baja probabilidad, hecho que se explica porque solo hay una manera de realizar estos números. Lo mismo ocurre con los números 3 y 18 en la suma de tres dados. En ambos casos, los números en el centro del gráfico tienen la mayor probabilidad de ocurrir, lo que se representa por una barra de mayor altura.

Supongamos que repetimos el experimento de los tres dados un gran número de veces y, cada vez, consignamos el valor de la suma. Podemos graficar la frecuencia con la que ocurrió cada número mediante un gráfico: ordenamos los posibles resultados en la horizontal y, sobre cada ubicación, dibujamos una barra de área proporcional a la frecuencia con la que ocurrió dicho resultado. Esto se conoce como **distribución empírica** o distribución de frecuencias. Los siguientes gráficos representan las distribuciones empíricas obtenidas al repetir el experimento de los tres dados 20, 100 y 10.000 veces. Observamos de inmediato que los gráficos se asemejan más al gráfico de la distribución de probabilidad teórica a medida que crece el número de repeticiones. ¡Esto es, otra vez, la **ley de los números grandes**! El gráfico de la distribución de frecuencias se acerca al gráfico de la distribución de probabilidad cuando se repite el experimento un gran número de veces.



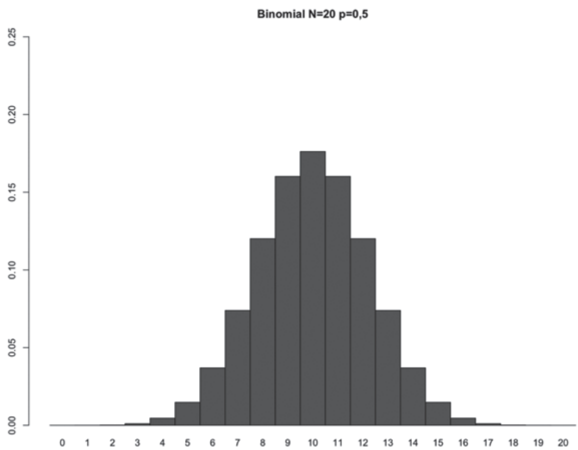




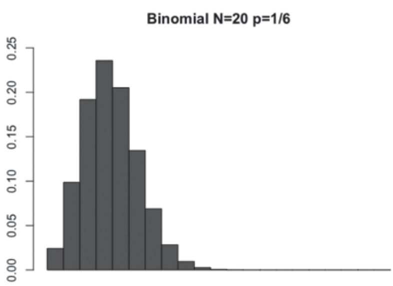
En el primer gráfico, vemos que algunos valores nunca aparecieron. Ningún lanzamiento resultó en los números 3, 4, 9, 16 y 18. Con 100 lanzamientos, los únicos valores faltantes son el 3 y el 18. Esto no es extraño, dado que son los números con menor probabilidad de ocurrir. En 10.000 lanzamientos, todos los números fueron logrados, con una frecuencia muy similar a la distribución de probabilidad obtenida de manera teórica.

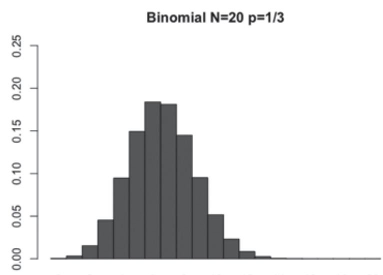
La distribución binomial

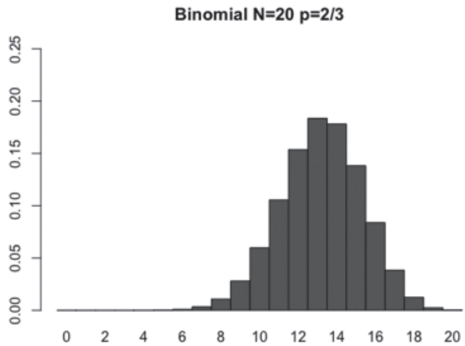
El siguiente gráfico representa una situación más compleja. En el experimento, se lanzan 20 monedas y se cuenta la cantidad de sellos obtenidos. Esto es un número entre 0 y 20. La distribución de probabilidad resultante se conoce como **distribución binomial** de parámetros **N** = 20 y **p** = ½.



La distribución binomial es más general. Representa la cantidad de éxitos obtenidos en un cierto número de experimentos idénticos que pueden ser exitosos con probabilidad **p** y fallar con probabilidad 1 – **p**. El valor **N** representa el número de experimentos repetidos y se relaciona con el tamaño del espacio muestra: los números entre 0 y **N**. Los siguientes gráficos representan distribuciones binomiales para 20 repeticiones de un mismo experimento —esto es, **N** = 20— para valores de **p** iguales a ⅙, ⅓ y ⅔ respectivamente.

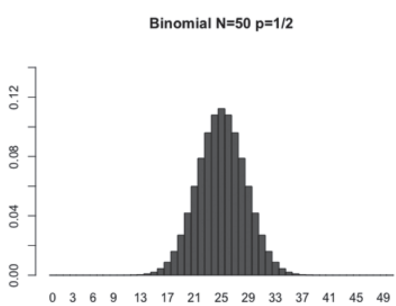


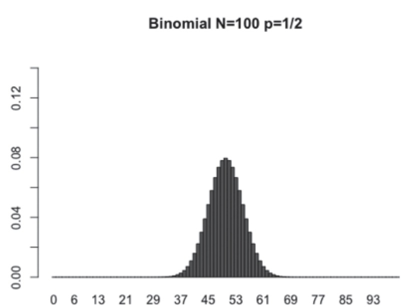


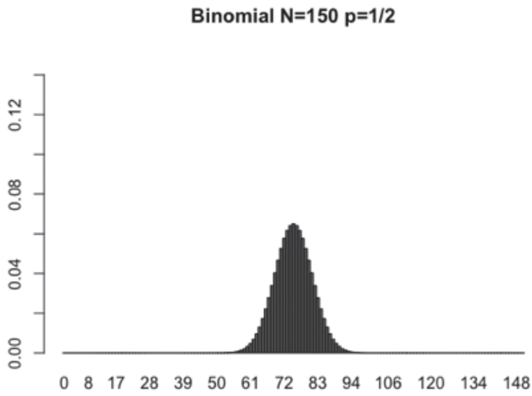


Vemos que, en cada caso, el gráfico se asemeja a una campana que se desplaza hacia la derecha a medida que aumenta el valor de **p**. Observamos que la cima de la campana se encuentra cerca del valor $\mathbf{N} \times \mathbf{p}$. Esto se relaciona con el promedio de la distribución, que estudiaremos en el capítulo 8. En el primer gráfico, $\mathbf{N} \times \mathbf{p} = 20 \times \frac{1}{6} \approx 3$. Un cálculo similar permite predecir la posición de la cima de la campana para los otros casos.

Finalmente, representamos en una misma escala los gráficos de distribuciones binomiales con $\mathbf{p} = \frac{1}{2}$, pero con cantidades de experimentos iguales a 50, 100 y 150 respectivamente. Vemos que el gráfico se vuelve más "pequeño" a medida que la cantidad de experimentos crece. Esto se debe a que las probabilidades de obtener cada uno de los números son cada vez más pequeñas. Dicho de otro modo, el espacio muestra crece con el valor **N**, y la distribución de probabilidad tiene que repartirse entre un mayor número de posibilidades.



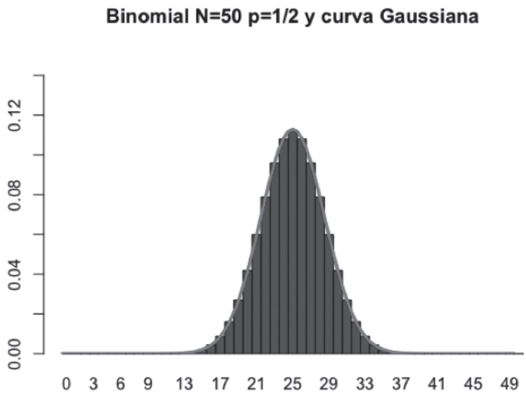




La tabla de Galton: la ley binomial se puede simular mediante un dispositivo conocido como tabla de Galton. Se trata de una tabla vertical con clavos dispuestos de manera alternada. Cuando se deja deslizar una gran cantidad de bolas diminutas por su superficie, estas golpean los clavos desordenadamente y se depositan en la base de la tabla, formando el gráfico de una distribución binomial.

La ley gaussiana

Los gráficos de distribuciones binomiales tienen la forma de una campana. Más abajo, hemos representado en un mismo gráfico una distribución binomial y una curva conocida como campana de Gauss o curva gaussiana. Esta es el análogo de un gráfico de distribución para una distribución de probabilidad que no es discreta —una distribución de probabilidad continua— conocida como la **ley gaussiana**. Las probabilidades de los eventos asociados a la ley gaussiana se entienden como el área contenida bajo porciones adecuadas de la campana, aunque esta interpretación es menos transparente que en el caso de distribuciones discretas. Ya aclararemos este concepto en nuestra discusión de variables aleatorias en el capítulo 8.



El gráfico anterior muestra que la distribución binomial se puede aproximar por una distribución continua, la ley gaussiana. Esta observación se debe originalmente a Abraham de Moivre y marca el descubrimiento del **teorema del límite central** en su versión más elemental. Como veremos más adelante, este teorema es muchísimo más general y se relaciona con la **ley de los errores** de Gauss, de donde la ley gaussiana recibe su nombre.

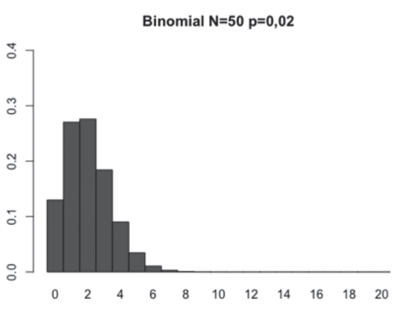
Dado que aparece en una gran variedad de contextos, la ley gaussiana se conoce también como "ley normal". Es omnipresente en la teoría de procesos estocásticos, que discutiremos en el capítulo 9. En ese contexto, caracteriza la posición de un objeto que se desplaza aleatoriamente o el precio de una acción en la bolsa. También aparece al caracterizar los errores de medición y la distribución de cantidades estadísticas provenientes de una multitud de situaciones. En general, caracteriza la ley de sistemas cuyo comportamiento se mantiene cercano al promedio. En efecto, una gran proporción del área bajo la campana de Gauss se concentra cerca de su cima. Las probabilidades representadas por las partes más achatadas de la campana son pequeñas.

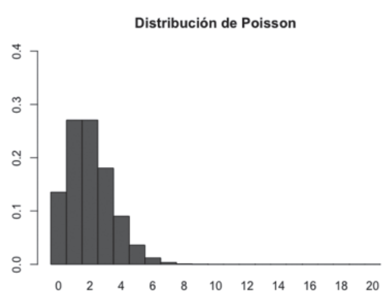
Existen distribuciones continuas muy distintas a la ley gaussiana que retratan situaciones en las que la mayoría de los eventos son cercanos al comportamiento promedio, pero en las que existen eventos extremadamente atípicos de probabilidad no despreciable. Estas leyes se conocen como **leyes de colas pesadas**, nombre que se refiere al hecho de que la curva que representa su gráfico no se vuelve pequeña tan rápidamente como la campana de Gauss. La magnitud de los sismos cae dentro de esta categoría: la mayoría de los sismos son benignos, pero, cada cierto tiempo, ocurre un sismo mayor.

Carl Friedrich Gauss (1777-1855) fue ciertamente una de las mentes más brillantes de la historia de la humanidad. Contribuyó de manera crucial a la matemática, la física y la astronomía, desde muy temprana edad. Gauss se negaba a publicar sus descubrimientos si los consideraba incompletos o insuficientemente buenos. En sus apuntes personales, escondió tesoros matemáticos muy adelantados para su época.

Poisson: la ley de lo improbable

Obtuvimos la ley gaussiana a partir de la ley binomial al considerar una gran cantidad de experimentos con una probabilidad razonable de éxito. Cuando esta probabilidad de éxito es muy pequeña, la situación es completamente diferente y lleva una distribución muy distinta a la gaussiana. Observamos a continuación el gráfico de una distribución binomial de parámetro **N** = 50 y **p** = 0,02 y la contraponemos con el gráfico de una nueva distribución conocida como la **distribución de Poisson**. La distribución de Poisson es una distribución discreta, aunque su espacio muestral se extiende a todos los números naturales y es, por lo tanto, infinito.





La distribución de Poisson se asocia a los eventos extremadamente raros. Fue observada por el economista y estadístico polaco Ladislaus Bortkiewicz (1868-1931) al analizar las muertes por pateaduras de caballo en el ejército prusiano a lo largo de 20 años. El matemático Siméon Poisson descubrió esta ley en 1837 al estudiar la distribución del número de convicciones equivocadas.

En un contexto general, la distribución de Poisson caracteriza el número de ocurrencia de eventos de baja probabilidad durante un intervalo de tiempo dado. Se relaciona con los tiempos de llegada de clientes en una cola o de buses en paraderos, con el decaimiento radiactivo de núcleos pesados y con las mutaciones genéticas. La comprobación de la naturaleza aleatoria de estas últimas se debe a los célebres experimentos de Salvador Luria y Max Delbrück de 1943. Esta ley también se aplica a contextos más insólitos, como el número de ataques de tiburones. Heurísticamente, la ley de Poisson aparece cuando se da un gran número de oportunidades, cada una de las cuales resulta con una probabilidad muy pequeña.

Siméon Poisson (1781-1840) fue un matemático francés recordado especialmente por sus contribuciones al análisis matemático y a sus aplicaciones a la física e ingeniería.

Si bien la ley de Poisson lleva su nombre, esta distribución apareció anteriormente en trabajos de Abraham de Moivre que datan de 1711. Sería justo renombrar esta ley como ley de De Moivre.

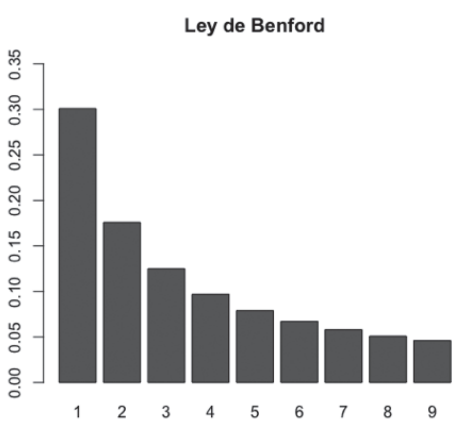
La ley de Benford

Supongamos que queremos inventar datos. Por ejemplo, tenemos una lista con datos incompletos que queremos rellenar, o crear transacciones ficticias en una declaración de impuestos. ¿Cómo generamos estos datos de manera creíble? Lo más probable es que recurramos a la distribución uniforme. Así, ningún número es sobrerrepresentado y el resultado es genuinamente aleatorio. La evidencia muestra que la realidad no es así. Los números aparecen según una ley propia conocida como la **ley de Benford**.

Empezamos nuestra discusión en un ámbito muy preciso: las tablas de logaritmos. Los logaritmos son omnipresentes en la física y es muy importante conocer sus valores numéricos. Antes del advenimiento de las calculadoras, calcular un logaritmo podía ser muy trabajoso. Para esto, existían libros con tablas de logaritmos. El astrónomo Simon Newcomb observó en 1881 que las primeras páginas de estos libros estaban más gastadas que las siguientes, mostrando que eran utilizadas más a menudo. Estas páginas contienen los números cuyo logaritmo empieza con el dígito 1, mostrando que la mayoría de las cantidades físicas de interés tienen un logaritmo que empieza con el dígito 1.

Benford analizó en 1937 más de 20.000 datos numéricos de distintos ámbitos y observó que el dígito 1 aparecía sistemáticamente en la primera posición con mayor frecuencia que los otros dígitos. Los datos analizados son tan diversos como números apareciendo en tapas de revistas, áreas de ríos, tablas matemáticas, tamaños de poblaciones y cantidades físicas. Realizó un análisis estadístico de la frecuencia de aparición de los distintos dígitos que aparecen en la primera posición de estos números, obteniendo una distribución de probabilidad conocida como la ley de Benford.

La ley de Benford es una simple distribución de probabilidad sobre los números del 1 al 9 que se grafica a continuación.



Vemos en particular que el número 1 tiene una probabilidad cercana a 30%, mucho mayor que la probabilidad 1/9 que le asignaría la probabilidad uniforme y mayor a la probabilidad asociada a los dígitos más grandes. Notamos que el verdadero valor correspondiendo al dígito 1 es, con mayor precisión, aproximadamente igual a 0,306, que, a su vez, es cercano al logaritmo de 2. La frecuencia del dígito 2 es, por otro lado, aproximadamente igual a la diferencia entre los logaritmos de 3 y de 2.

Esto nos da una herramienta para testear la veracidad de ciertos datos que sospechamos fraudulentos. Datos reales, a menudo, deberían parecerse a la ley de Benford más que a la distribución uniforme.

Podemos argumentar que los datos que utilizamos están calibrados según algún sistema de unidades más o menos arbitrario —como kilómetros o millas— y que la ley de Benford pudiera ser una particularidad de nuestro sistema métrico. Resulta que la altura de los mayores edificios del mundo sigue la ley de Benford sin importar si son medidos en metros o en pies.

Este es un ejemplo de universalidad. Algunas leyes de probabilidad aparecen en contextos dispares que comparten algunas vagas características. La ley gaussiana es un ejemplo de universalidad. La ley de Benford pareciera ser la norma para grandes cantidades de datos de la vida cotidiana cuyo tamaño varía apreciablemente, es decir, bases de datos con datos pequeños, moderados, grandes y muy grandes.

Algunos datos, sin embargo, distan de seguir la ley de Benford, como cantidades provenientes de manuales de ingeniería o ciertas constantes físicas.

LA PROBABILIDAD CONDICIONAL

La probabilidad condicional es un concepto central de la teoría de probabilidades. Básicamente, consiste en restringir el universo de posibilidades, creando así un nuevo espacio de probabilidad más pequeño.

Varias paradojas nacen de un entendimiento erróneo de la probabilidad condicional. La interpretación correcta de algunos datos —como los resultados de pruebas médicas— es un asunto de probabilidad condicional.

La paradoja de las dos hijas

Para empezar, discutiremos algunas variantes de una misma historia. La paradoja de las dos hijas es un problema clásico de probabilidad introducido en 1959 por el matemático Martin Gardner, uno de los pioneros de la divulgación matemática.

> *Una pareja tiene dos hijos/hijas. ¿Cuál es la probabilidad de que ambas sean mujeres?*

Es sabido que la probabilidad de que nazca una mujer es levemente más elevada que la probabilidad de que nazca un hombre. Sin embargo, para simplificar el problema, supondremos que ambas probabilidades son iguales a ½. Listemos todas las posibilidades para el sexo de dos hijos:

Primer hijo/hija	Segundo hijo/hija
Mujer	Mujer
Mujer	Hombre
Hombre	Mujer
Hombre	Hombre

Estas cuatro alternativas tienen todas la misma probabilidad, que debe, por lo tanto, ser igual a ¼. La probabilidad de que ambos hijos/hijas sean mujeres es igual a ¼.

Veamos cómo cierta información adicional puede afectar el problema. Consideremos la siguiente situación:

Una pareja tiene dos hijos/hijas. Sabemos que la menor es mujer. ¿Cuál es la probabilidad de que ambas sean mujeres?

El problema equivale a determinar el sexo del segundo hijo. Ahora, el sexo del primer y del segundo hijo son cantidades independientes. Por lo tanto, sin importar el sexo del primer hijo, la probabilidad de que el segundo sea mujer es ½. La información adicional no modifica el problema.

La siguiente variante es mucho más interesante:

Una pareja tiene dos hijos/hijas. Sabemos que al menos uno de ellos es mujer. ¿Cuál es la probabilidad de que ambas sean mujeres?

Es tentador razonar de la manera siguiente: el sexo del primer y del segundo hijo son cantidades independientes. Por lo tanto, sin importar que sabemos que ya hay una hija, la probabilidad de que el otro sea mujer debe ser ½. Esto es incorrecto.

Para obtener la respuesta correcta, tenemos que volver a lo básico y calcular el cociente entre el número de posibilidades favorables y el número de posibilidades totales. Aquí es donde la información adicional entra en juego. Sabemos que hay al menos una hija, por lo tanto, el número de posibilidades totales se reduce a tres: las tres primeras posibilidades listadas en la tabla de más arriba. Estas tienen cada una la misma probabilidad de ocurrir. De estas tres, solo la primera corresponde a una situación favorable. Por lo tanto, la probabilidad es ⅓.

Este es un ejemplo de probabilidad condicional: el conocimiento de cierta información adicional nos permite restringir los casos posibles y trabajar con un espacio muestral más pequeño.

> **Martin Gardner (1914-2010)** fue un prolífico autor de obras de divulgación científica. Escribió durante 25 años una columna de juegos matemáticos en los que presentaba la matemática de forma accesible y entretenida. Fue un pionero de la divulgación matemática y de la matemática recreacional, y un ingenioso creador de puzles mentales como la paradoja de las dos hijas. Fue también un gran admirador de la obra del eximio matemático inglés Charles Dodgson, mejor conocido como Lewis Carroll y como autor de *Alicia en el País de las Maravillas*. Gardner publicó una versión anotada de *Alicia*.

La probabilidad condicional

Aprendimos a calcular la probabilidad de un evento calculando el cociente del número de casos favorables y del número de casos totales. Por ejemplo, para calcular la probabilidad de que el número arrojado por un dado sea par, contamos los casos favorables —tres casos exactamente— y los casos totales —seis casos en total—. La probabilidad es, por lo tanto, 3/6 = ½ = 0,5.

Ahora, ¿cuál es la probabilidad de que el número arrojado por un dado sea par si sabemos además que el resultado es estrictamente mayor que 3? El número de casos posibles se redujo: ahora, solo tenemos que tomar en cuenta los resultados 4, 5 y 6. De estos tres valores, solo el 4 y el 6 son números pares. Luego, la probabilidad buscada es ⅔ ≈ 0,66. Matemáticamente, decimos que la probabilidad condicional de que el número arrojado por el dado sea par dado que este número es estrictamente mayor que 3 es ⅔.

En general, pensemos que tenemos un espacio muestral sobre el cual se define una medida de probabilidad mediante algún experimento. Luego, incorporamos información al problema: sabemos que el resultado del experimento cae dentro de un subconjunto del espacio muestral.

Llamaremos **A** a este subconjunto. Se trata de calcular la probabilidad de ocurrencia de alguna de las posibilidades tomando en cuenta que tiene que estar dentro de **A**: esto es, su probabilidad condicional dado **A**.

Si una posibilidad no cae dentro de **A**, su probabilidad condicional dado **A** es igual a 0. Pero, si cae dentro de **A**, tenemos que tomar en cuenta su probabilidad de ocurrencia. Calcular esta nueva probabilidad es sencillo. La probabilidad condicional dado **A** es una medida de probabilidad, por lo que debe asignarle probabilidad igual a 1 a su espacio muestral. Es decir, la probabilidad condicional de **A** dado **A** es igual a 1. Para calcular la probabilidad condicional de un elemento de **A**, basta con calcular su probabilidad original y dividirla por la probabilidad de **A**.

No tendremos que recurrir a estos argumentos abstractos. Para todos los efectos, cuando sea necesario calcular una probabilidad condicional, solo seremos cautelosos en identificar el nuevo espacio muestral y calcular el cociente entre casos favorables y casos totales.

Falsos positivos

El siguiente ejemplo nos muestra cómo se deben interpretar los resultados de pruebas médicas:

> *Una cierta enfermedad tiene una prevalencia de 1/1.000, es decir, una de cada mil personas tiene la enfermedad en promedio. Existe una prueba médica que siempre arroja un resultado positivo cuando una persona tiene la enfermedad. Sin embargo, la probabilidad de un falso positivo es de 5%: la probabilidad de que un individuo sano tenga un resultado positivo es 5%. Dicho de otro modo, la probabilidad condicional de que el test dé positivo dado que el individuo es sano es de 5%.*

La pregunta que intentaremos contestar es:

> *Si una persona obtiene un resultado positivo, ¿cuál es la probabilidad de que tenga la enfermedad?*

Afortunadamente, esta probabilidad no es 95%. Tenemos que leer los datos con cuidado. 5% corresponde a la probabilidad condicional de que el test sea positivo dado que el individuo está sano. Por lo tanto, 95% corresponde a la probabilidad condicional de que el test sea negativo dado que el individuo está sano. Nos interesa la probabilidad condicional de que el individuo esté enfermo dado que el test es positivo.

Sabemos que solo una persona de cada mil tiene la enfermedad. También sabemos que 5% de estas mil personas, es decir, 50 personas, tendrán un test positivo a pesar de estar sanas. Es decir, el test resultará positivo 1 + 50 = 51 veces. Si el test de una persona es positivo, es muchísimo más probable que se trate de uno de los falsos positivos. Luego, la probabilidad buscada debe ser bastante pequeña.

Volviendo a nuestro cálculo: nuestro espacio de probabilidad se reduce a los casos en que el test es positivo: 51 casos. De estos 51 casos, solo uno corresponde a una persona enferma. Es decir, la probabilidad condicional de que una persona tenga la enfermedad dado que el test es positivo es 1/51, un poco menos de 2%.

De modo que, aun si el test resulta positivo, no hay motivo para entrar en pánico. Hay una probabilidad sobre 98% de estar sano a pesar de todo.

Dejamos una pregunta abierta para el lector: ¿cuál es la probabilidad de que una persona esté enferma si el resultado del test fue positivo en dos oportunidades?

La regla de Bayes: en situaciones más complejas, los problemas de este tipo se resuelven usando el teorema de Bayes o ley de Bayes. Esta ley permite "invertir" las probabilidades condicionales. Por ejemplo, conociendo la probabilidad de que una prueba médica arroje un resultado positivo dado que una persona está enferma, podemos inferir la probabilidad de que una persona esté enferma dado que el test resultó positivo, tal como procedimos en el ejemplo anterior. Nuestro procedimiento fue elemental. El problema se complica si incluimos la posibilidad de falsos negativos.

Thomas Bayes (1701-1761) fue un estadístico inglés. Una de las corrientes principales de la estadística lleva su nombre: la estadística bayesiana, que incorpora la noción subjetiva de probabilidad que mencionamos en la introducción.

Taxista en fuga

El siguiente problema se encuentra en un artículo de Amos Tverski, de la Universidad de Stanford, y Daniel Kahneman, de la Universidad de British Columbia. A grandes rasgos, Tverski y Kahneman estudiaban el comportamiento de individuos frente a situaciones que involucran probabilidades y ponen de manifiesto que nuestra intuición nos puede llevar a conclusiones equivocadas.

El problema es el siguiente:

> *En una ciudad, operan dos compañías de taxis: los taxis verdes y los taxis azules. Se sabe que 85% de los taxis son verdes y 15% son azules.*
>
> *Una noche, un taxista comete una imprudencia, provoca un grave accidente de tránsito y se da a la fuga.*
>
> *¿Cuál es la probabilidad de que se trate de un taxi azul?*

La respuesta a este problema es sencilla: esta probabilidad es igual a 15%. Agreguemos nueva información:

> *Un único testigo identifica el taxi como un taxi azul.*
>
> *¿Cuál es la probabilidad de que el taxi involucrado en el accidente sea un taxi azul?*

Curiosamente, el nuevo dato no nos permite llegar a una conclusión. Sigamos con la historia:

> *El juez encargado del caso diseña el siguiente experimento para determinar la confiabilidad del testigo: bajo las mismas circunstancias que el accidente, es decir, en el mismo lugar, a la misma hora y con las mismas condiciones de luminosidad, se le pide al testigo de identificar taxis como verdes o azules. El testigo acierta 80% de las veces.*
>
> *¿Cuál es la probabilidad de que el taxi involucrado en el accidente sea realmente un taxi azul?*

A primera vista, pareciera muy probable que el taxi sea efectivamente azul, ya que la tasa de éxito del testigo es altísima.

Para resolver este problema, tenemos que identificar correctamente algunas probabilidades condicionales. Lo que intentamos averiguar es la probabilidad condicional de que el taxi sea azul dado que el testigo lo identificó como azul. Nuevamente, se trata de contar correctamente.

Para simplificar el problema, digamos que hay 100 taxis en la ciudad. 85 de ellos son verdes y 15 son azules. El número de casos favorables en este problema es el número de taxis azules que el testigo identifica correctamente como azules. El número de casos totales es el número total de taxis que el testigo identifica como azules. La probabilidad buscada es el cociente de estos dos números.

Contemos primero los casos favorables: de los 15 taxis azules, solo el 80% es identificado correctamente como azul, es decir, 12 taxis.

Los casos totales son de dos tipos: los taxis azules identificados como azules y los taxis verdes identificados erróneamente como azules. Ya sabemos cuántos son del primer tipo: 12. De los 85 taxis verdes, el 20% es identificado incorrectamente como azul: 17 taxis. En resumen, el número total de taxis identificados como azules es 12 + 17 = 29.

La probabilidad buscada es entonces igual a 12/29: un poco más de 41%. Es decir, aunque el taxi haya sido identificado como azul, hay una probabilidad de más de 50% de que se trate en realidad de un taxi verde.

TRES PUZLES CON EL NÚMERO 3

El mundo de las probabilidades abunda en puzles y paradojas que desafían la intuición. Daremos tres ejemplos que ya pertenecen al folclore de esta disciplina.

El juego de Monty Hall

A pesar de estar rodeados de azar, pensar probabilísticamente no es fácil. Por lo mismo, es muy fácil engañar a la intuición. El mejor ejemplo de esto es un peculiar concurso de la televisión norteamericana que hoy en día lleva el nombre del anfitrión del show: el problema de Monty Hall. El problema es muy fácil de formular, pero la respuesta es tan sorprendente que la mayoría de las personas —incluso los matemáticos— tarda mucho en convencerse de que es correcta. Aquí va:

> *Hay tres puertas cerradas en un set de televisión. Tras una de ellas, hay un auto cero kilómetros; tras las dos otras, una cabra. Solo el anfitrión sabe dónde está el auto. El concursante debe elegir una puerta. Si adivina qué puerta esconde el auto, se lo lleva de premio.*
>
> *¡Ahora es cuando el juego se vuelve interesante! El concursante elige una puerta, digamos la puerta número 1. Acto seguido, el anfitrión abre una de las otras puertas, tras la cual sabe que hay una cabra —digamos la número 2— y le ofrece al concursante la posibilidad de cambiar su elección original y elegir la puerta número 3.*
>
> *¿Cuál es la mejor estrategia? ¿Cambiar de puerta o mantener la elección original?*

La primera respuesta que viene a la mente es la siguiente: una vez que eliminamos la puerta número 2, solo hay dos posibilidades para la ubicación del auto: la puerta número 1 y la puerta número 2. Luego, la probabilidad de que el auto se encuentre tras la puerta número 1 es ½ y lo mismo para la probabilidad de que el auto se encuentre tras la puerta número 3. Por lo tanto, quedarse con la puerta original o cambiar es irrelevante.

Esta respuesta es la que gran parte de nosotros daría. Es extremadamente natural. Desgraciadamente, es incorrecta.

Para encontrar la solución correcta, una vez más, tenemos que contar las distintas posibilidades cuidadosamente. Al inicio, la probabilidad de que el auto se encuentre tras cualquiera de las tres puertas es ⅓. En este sentido, la elección original de la puerta es irrelevante, por lo que supondremos que el concursante elige la puerta número 1.

La siguiente tabla ilustra todos los posibles casos:

Puerta 1	Puerta 2	Puerta 3	Si se queda	Si cambia
Auto	Cabra	Cabra	Gana	Pierde
Cabra	Auto	Cabra	Pierde	Gana
Cabra	Cabra	Auto	Pierde	Gana

En el primer caso, el auto se encuentra tras la puerta número 1. El anfitrión abrirá cualquiera de las otras dos puertas. Si el concursante se queda, gana; si no, pierde.

En el segundo caso, luego de que el concursante elige la puerta número 1, el anfitrión solo puede abrir la puerta número 3, ya que el auto se encuentra tras la puerta número 2. En este caso, si el concursante se queda con la puerta número 1, pierde.

El tercer caso es similar. El anfitrión solo puede abrir la puerta número 2 y el concursante debería cambiarse a la puerta número 3 para ganar.

Luego, en dos de los tres casos, cambiarse de puerta es la estrategia ganadora. ¡Por lo tanto, la probabilidad de ganar cambiándose de puerta es ⅔! Conclusión: cambiarse de puerta aumenta nuestras chances de ganar.

Otra forma de razonar es la siguiente: originalmente, la probabilidad de que el auto esté tras la puerta número 1 es ⅓. Esta probabilidad no puede cambiar por el mero hecho de abrir una de las otras dos puertas. Por lo tanto, como las probabilidades deben sumar 1, la probabilidad de que el auto se encuentre tras la otra puerta sin abrir debe ser ⅔.

A pesar de comprobar que la mayoría de los concursantes que cambian de puerta ganan, muchos televidentes no se convencieron de la respuesta correcta y siguieron creyendo que la probabilidad de ganar quedándose con la puerta original es igual a ½. Es decir, incluso observando la frecuencia de ganadores y perdedores, que estaba claramente a favor de aquellos que cambiaban de puerta, no fueron capaces de volver a estimar esta probabilidad. Así funciona la mente humana: no somos tan buenos para las probabilidades.

Dato curioso: se han conducido experimentos del tipo Monty Hall con palomas. Las palomas rápidamente aprendieron que les convenía cambiar de puerta. Algunos animales son mejores que nosotros en lidiar con situaciones inciertas. Aprenden a reaccionar basados en su experiencia, es decir, estimando las frecuencias observadas. En nuestro caso, es posible que nuestra poca acertada intuición no nos permita actualizar nuestra apreciación subjetiva de la probabilidad de un evento, llevándonos a errores repetidos.

La paradoja de las tres cajas

El problema de Monty Hall es muy similar a un problema mucho más antiguo conocido como la paradoja de Bertrand o la paradoja de las tres cajas.

> *Hay tres cajas. La primera contiene dos monedas de plata; la segunda, dos monedas de oro; la tercera, una moneda de plata y una moneda de oro.*
>
> *Se elige una caja al azar y, luego, se elige al azar una de las dos monedas que contiene esa caja. Resulta ser una moneda de oro.*
>
> *¿Cuál es la probabilidad de que la segunda moneda también sea de oro?*

Tal como en el problema de Monty Hall, hay solo dos posibilidades para la segunda moneda: plata u oro. Por lo tanto, la probabilidad de que la segunda moneda también sea una moneda de oro es ½. Nuevamente, esta respuesta es incorrecta y, nuevamente, basta con contar cuidadosamente las distintas posibilidades para calcular correctamente esta probabilidad. De nuevo, esto reduce a un cálculo cuidadoso de probabilidades condicionales.

Al elegir primero una caja y luego las monedas una tras otra, hay seis posibilidades, retratadas en la siguiente tabla:

Caja	Primera moneda	Segunda moneda
Número 1	Plata	Plata
	Plata	Plata
Número 2	Oro	Oro
	Oro	Oro
Número 3	Plata	Oro
	Oro	Plata

Hay solo tres posibilidades para que la primera moneda sea una moneda de oro: los casos 3, 4 y 6. De estos tres casos, solo los casos 3 y 4 brindan una segunda moneda de oro. Luego, dos de los tres casos posibles son favorables y la probabilidad buscada es $2/3$.

La paradoja de los tres prisioneros

La paradoja de los tres prisioneros funciona de manera parecida. Existen múltiples variantes, de las cuales presentaremos solo una. La solución se deja como ejercicio para el lector.

> *Hay tres prisioneros: Alberto, Bernardo y Claudio. Uno de ellos, escogido al azar, será ejecutado y los otros serán liberados. El guardia escuchó que Bernardo iba a ser liberado y esta es toda la información de la que dispone. Alberto, que no sabe si morirá, quiere enviarle una carta a su familia. Como solo un prisionero será ejecutado, o bien Bernardo o bien Claudio (o ambos) saldrán libres y podrán llevar la carta. Alberto le pide al guardia que, de saberlo, le indique el nombre de uno de los prisioneros que será liberado.*
>
> *El guardia razona de la manera siguiente: si doy el nombre de Bernardo, la probabilidad de que Alberto sea ejecutado sube de $1/3$ a $1/2$. Por lo tanto, mejor no digo nada.*

Finalmente, el guardia accede y da el nombre de Bernardo. Claudio escuchó el diálogo y piensa: ahora, la probabilidad de que yo sea ejecutado subió de ⅓ a ⅔.

¿Quién tiene la razón? ¿Claudio o el guardia?

VARIABLES ALEATORIAS Y ESPERANZAS

El título de este capítulo es engañoso. Por *esperanza* no nos referimos a ese sentimiento hermoso que nos ilumina el futuro, sino que a un objeto central de la teoría de probabilidades.

Cuando lidiamos con eventos azarosos, calculamos probabilidades para cuantificar las chances de ocurrencia de cada posibilidad. Cuando lidiamos con cantidades azarosas, nos interesa calcular su **promedio**. En teoría de probabilidades, un valor azaroso se conoce como **variable aleatoria**.

Por ejemplo, en un juego de cara o sello, A y B apuestan cada uno 10 doblones. Si sale cara, A se lleva los 20 doblones; si sale sello, B se los lleva. Con probabilidad ½, A se llevará 10 doblones adicionales a su apuesta inicial, mientras que, con probabilidad ½, perderá los 10 doblones que apostó. Dicho de otro modo, A gana 10 doblones con probabilidad ½ y pierde 10 doblones con probabilidad ½. Por lo tanto, *en promedio*, no gana nada y no pierde nada. Así, al final del juego, A tendrá, *en promedio*, 10 doblones.

Recordemos que en el juego de dados del Caballero de Méré el jugador tiene una probabilidad levemente mayor que ½ de ganar. Si el jugador y la casa apuestan 10 doblones, el jugador ganará 10 doblones con probabilidad levemente mayor que ½ y perderá 10 doblones con probabilidad levemente menor que ½. Por lo tanto, su *ganancia promedio* será positiva. La casa, por otro lado, tiene una "ganancia" promedio negativa, es decir, en promedio, pierde dinero.

En la teoría de probabilidades, esta ganancia promedio se conoce como **valor esperado** o **esperanza**. Los términos *promedio*, *valor esperado* y *esperanza* son sinónimos. Cuando nos referimos a algo que ocurre "en promedio", nos referimos a alguna propiedad de la esperanza.

Por ejemplo, a menudo diremos que una cantidad es cero en promedio. Significa que la variable aleatoria correspondiente tiene esperanza igual a cero. Ya veremos cómo se calculan las esperanzas en casos concretos.

La paradoja de la corbata

Como todos los conceptos de la teoría de probabilidades, el promedio o esperanza da lugar a paradojas extraordinarias. La primera paradoja que enunciaremos se conoce como la ***paradoja de la corbata***.

> *Dos hombres, A y B, reciben cada uno una corbata para Navidad de parte de sus esposas y debaten acerca de quién recibió la corbata más cara. Para zanjar la discusión, deciden preguntar a sus respectivas esposas el precio de sus respectivas corbatas, no sin antes involucrarse en la siguiente apuesta absurda: ¡el que recibió la corbata más cara se la regala al otro!*
>
> *El razonamiento de A es el siguiente: tengo una probabilidad ½ de tener la corbata más cara; en este caso, pierdo el valor de mi corbata. Por otro lado, tengo una probabilidad ½ de tener la corbata más barata; en este caso, gano un valor mayor al valor de mi corbata. Luego, en promedio, tengo una ganancia positiva. Por supuesto, B razona de la misma manera.*
>
> *Por lo tanto, pareciera que el juego es favorable simultáneamente para ambos jugadores.*
>
> *¿Cómo es esto posible?*

Por supuesto, hay un error en este razonamiento. La paradoja desaparecerá una vez que tengamos un concepto más claro de promedio.

El promedio o valor esperado

Una variable aleatoria es una cantidad aleatoria. Por ejemplo, el resultado de lanzar un dado es una variable aleatoria que toma valores entre 1 y 6. Esta variable aleatoria asume cualquiera de estos valores con probabilidad ⅙. En cara o sello, podemos codificar los resultados a través de números. Podemos asignarle a cara el número 1 y a sello, el número 0. Así, el resultado de tirar una moneda es una variable aleatoria que puede tomar los valores 0 o 1. En probabilidades, una variable aleatoria que puede asumir solo dos valores distintos se conoce como una **variable aleatoria de Bernoulli**.

Pensemos ahora que apostamos 10 doblones en un juego de cara o sello. La ganancia es una cantidad aleatoria o variable aleatoria. Toma el valor 10 cuando ganamos, lo que ocurre con probabilidad ½, y el valor –10 cuando perdemos, también con probabilidad ½. Puede resultar extraño hablar de ganancias negativas, pero es matemáticamente conveniente tratar las ganancias y las pérdidas de manera unificada. El promedio o esperanza de esta variable aleatoria es igual a 0, aunque esta no asuma nunca el valor 0. Esto puede ilustrarse de la siguiente manera: supongamos que tenemos una balanza sobre la cual disponemos los doblones. A la derecha ponemos los 10 doblones que ganaríamos con probabilidad ½ y a la izquierda los 10 doblones que perderíamos con probabilidad ½. La balanza no se inclina en ninguna dirección. El juego es balanceado de algún modo. Es un juego justo.

El promedio, valor esperado o esperanza nos da una medida de las cantidades aleatorias. Para calcularla, tenemos que contabilizar todos sus valores posibles y ponderarlos por su respectiva probabilidad.

Kolmogorov y la noción de promedio: Kolmogorov no solo formuló los axiomas de la probabilidad, sino que también caracterizó el concepto de promedio en base a algunas propiedades simples. Su genial abstracción de la noción de promedio cubre, entre otros, las versiones más familiares, que son la media aritmética, la media geométrica y la media armónica.

Calculemos algunas esperanzas. Volvamos al cara o sello codificado por 0 y 1. La variable aleatoria que nos da el resultado del juego toma el valor 1 con probabilidad ½ y el valor 0 con probabilidad ½. La esperanza es igual a $1 \times ½ + 0 \times ½ = ½$. Aun si nuestra variable aleatoria nunca asume el valor ½, su esperanza o promedio es igual a ½.

¿Cuál es la esperanza del resultado obtenido al tirar un dado? Los valores posibles de esta variable aleatoria son los números del 1 al 6, cada uno obtenido con probabilidad ⅙. La esperanza es:

$$1\times\tfrac{1}{6}+2\times\tfrac{1}{6}+3\times\tfrac{1}{6}+4\times\tfrac{1}{6}+5\times\tfrac{1}{6}+6\times\tfrac{1}{6}=21/6=3{,}5$$

Este valor se encuentra entre el 3 y el 4. Esto es razonable: hay tres números menores que 3,5 y tres números mayores que este valor. La esperanza los separa equitativamente.

Cuándo y cuánto apostar: en un juego justo, la esperanza de la ganancia es 0. En cara o sello, es claro que ambas partes tienen que apostar la misma cantidad. Hay circunstancias más complejas en las que es posible inclinar esta esperanza a nuestro favor.

En el póker, podemos calcular en cualquier instante dado nuestra probabilidad de ganar y tenemos que apostar para seguir en el juego. Esto nos permite calcular la máxima apuesta que podemos pagar para tener una esperanza positiva en nuestras ganancias. Si nuestra probabilidad de ganar es $\mathbf{p}$, nuestra probabilidad de perder es $1 - \mathbf{p}$. El valor de nuestra apuesta, que representaremos por $\mathbf{a}$, es la cantidad de dinero que podríamos perder. El valor del pozo, denotado por $\mathbf{b}$, es el dinero que podríamos ganar. Es decir, ganamos $\mathbf{b}$ con probabilidad $\mathbf{p}$ y perdemos $\mathbf{a}$ con probabilidad $1 - \mathbf{p}$. Nuestra ganancia promedio es por lo tanto igual a $\mathbf{b} \times \mathbf{p} - \mathbf{a} \times (1 - \mathbf{p})$. Tenemos que asegurarnos de que esta cantidad sea positiva. Observemos que esta cantidad decrece cuando $\mathbf{a}$ crece y, si $\mathbf{a}$ es demasiado grande, se vuelve negativa; es decir, perdemos dinero en promedio. Si se nos exige una apuesta demasiado grande para seguir jugando, lo más sensato es abandonar.

Existe una versión de la ley de los números grandes para variables aleatorias. Supongamos que nuestra variable aleatoria proviene de algún experimento. Cada vez que repitamos el experimento, esta tomará un valor distinto. Al repetir este experimento varias veces, podemos consignar estos valores y promediarlos; nos referimos al promedio aritmético usual, no a la esperanza. La ley de los números grandes nos dice que, cuando realicemos un gran número de experimentos, este promedio será muy cercano a la esperanza de la variable aleatoria. Como en el caso de la ley de los números grandes para distribuciones de probabilidad, el promedio aritmético de las variables aleatorias será en general distinto de la esperanza. Lo importante es que la diferencia entre estos valores es pequeña comparada con la cantidad total de variables aleatorias. Esta diferencia se caracterizará en el teorema central del límite.

S. R. S. Varadhan y la teoría de grandes desvíos: hay muchos mitos acerca del probabilista indio S. R. S. Varadhan. Se dice de él que resolvió un problema muy difícil subiendo unos pisos en ascensor y que ideó la teoría de grandes desvíos quedándose sentado durante algunas horas tras una pregunta de un colega.

La ley de los números grandes asegura que el promedio aritmético de la suma de una gran cantidad de variables aleatorias idénticas se acerca a la esperanza común de estas variables aleatorias. Hay, sin embargo, una probabilidad pequeña y evanescente de que estas cantidades difieran de manera significativa. La teoría de grandes desvíos caracteriza estos eventos atípicos. Nos permite estimar la probabilidad de estos eventos de "grandes desvíos" con respecto del promedio. En situaciones más complejas que las sumas de variables aleatorias, incluso entrega información sobre las estrategias seguidas por los sistemas para vencer la ley de los números grandes.

Resolviendo la paradoja de la corbata

Ahora que tenemos un conocimiento más acabado de la noción de promedio, podemos resolver la paradoja de la corbata.

Supongamos, para fijar las ideas, que la corbata más cara cuesta 20 doblones y la otra, 10 doblones.

Con probabilidad ½, A tiene la corbata más barata y gana 20 doblones. Por otro lado, con probabilidad ½, A tiene la corbata más cara y, por lo tanto, pierde 20 doblones. Luego, la ganancia promedio de A es $20 \times ½ - 20 \times ½ = 0$. Del mismo modo, la ganancia promedio de B es 0.

Era absurdo pensar que A y B tienen ambos una ganancia promedio positiva. La realidad es que ambos tienen una ganancia promedio igual a 0, es decir, el juego es justo.

Un cálculo más general: en el cálculo anterior, le asignamos un valor a cada corbata para fijar las ideas. El resultado del cálculo no depende del valor preciso de las corbatas. Digamos que la corbata más barata cuesta **X**. La más cara cuesta, por lo tanto, 2**X**. El jugador A tiene la corbata de valor **X** con probabilidad ½, en cuyo caso gana 2**X**. Asimismo, tiene la corbata de valor 2**X** con probabilidad ½. En este caso, pierde 2**X**. La ganancia esperada es entonces igual a $2\mathbf{X} \times ½ - 2\mathbf{X} \times ½ = 0$.

El dilema del prisionero

El dilema del prisionero es un ejemplo clásico de teoría de juegos en el que la suerte de dos "jugadores" —los prisioneros— depende de su disposición a cooperar o traicionar. Consideraremos el problema original y una variante que permite un análisis probabilístico más interesante.

La primera situación es como sigue:

> *Un rico señor es asesinado en su hacienda. Las cámaras de seguridad registran que dos personas, A y B, ingresaron ilícitamente a la propiedad la noche de los hechos. Por supuesto, ambos son sospechosos del asesinato, pero también podría tratarse simplemente de dos ladrones que se encontraron en el lugar equivocado en el momento equivocado. La sentencia por ingresar ilegalmente a la propiedad es de un año de cárcel, mientras que la sentencia por asesinato es de 14 años.*
>
> *A y B son llevados a la comisaría, aislados el uno del otro e interrogados separadamente. Ambos arriesgan al menos una condena de un año por violación de propiedad privada.*
>
> *El detective le propone a A el siguiente trato: si traiciona a su compañero, A queda libre y B es condenado a 14 años por asesinato. Por supuesto, le propone a B el mismo trato. Sin embargo, les advierte lo siguiente: si ambos traicionan, ambos serán condenados a 10 años de cárcel por complicidad, pues, si bien la autoría del asesinato resultaría dudosa en este caso, esto demostraría que ambos están involucrados en el asesinato de alguna manera.*

La tabla siguiente ilustra las diferentes posibilidades:

	A coopera	A traiciona
B coopera	A: 1 año B: 1 año	A: 0 años B: 14 años
B traiciona	A: 14 años B: 0 años	A: 10 años B: 10 años

El razonamiento es sencillo. Cada uno debe traicionar. Si A traiciona, se arriesga a lo más a una condena de 10 años y, en el mejor de los casos, sale libre. Por otro lado, si coopera, se arriesga al menos a un año de cárcel y, en el peor de los casos, a 14 años de presidio.

Podemos incluir algo de probabilidades en el razonamiento. Supongamos que A confía plenamente en su compañero, es decir, la probabilidad de que B traicione es igual a 0. Luego, al traicionar, A saldrá libre con plena certeza, mientras que, si coopera, cumplirá un año de condena. Si, por otro lado, A considera que la probabilidad de que su compañero lo traicione es igual a 1, A cumplirá 10 años de cárcel al traicionar y 14 al cooperar. La decisión es clara. Ahora, A puede considerar que la probabilidad de ser traicionado es cercana a ½. En este caso, al cooperar, A cumplirá una condena de un año con probabilidad ½ y una condena de 14 años con probabilidad ½, es decir, su condena promedio será igual a $½ \times 1 + ½ \times 14 = 7{,}5$. Por otro lado, si A traiciona, cumplirá una condena de 0 años con probabilidad ½ y una condena de 10 años con probabilidad ½, es decir, en este escenario, su condena promedio será igual a $0 \times ½ + 10 \times ½ = 5$. Otra vez, la situación es clara.

Se puede ver que, independientemente de la confianza de A en la integridad de su compañero, traicionar es la opción ganadora. Notemos, sin embargo, que cooperar minimiza la suma de las condenas.

Compliquemos la situación:

> *Supongamos que todo sigue igual salvo que, si A traiciona y B coopera, entonces B es condenado a 14 años por asesinato y A, a cinco años por complicidad; y viceversa.*

La tabla siguiente ilustra la nueva situación:

	A coopera	A traiciona
B coopera	A: 1 año B: 1 año	A: 5 años B: 14 años
B traiciona	A: 14 años B: 5 años	A: 10 años B: 10 años

La situación es mucho más compleja. Si A confía en que su compañero cooperará, le conviene cooperar. Sin embargo, si cree que su compañero lo traicionará, entonces es mejor traicionar también.

¿Qué pasa si A considera que la probabilidad de que su compañero lo traicione es igual a ½? En este caso, al cooperar, cumplirá una condena de un año con probabilidad ½ y una condena de 14 años con probabilidad ½. Es decir, su condena promedio será igual a $1 \times ½ + 14 \times ½ = 7{,}5$. Por otro lado, en caso de traicionar, cumplirá una condena de cinco años con probabilidad ½ y una condena de 10 años con probabilidad ½. En este caso, su condena promedio será igual a $5 \times ½ + 10 \times ½ = 7{,}5$. En promedio, ambas estrategias son equivalentes.

¿Qué hacer entonces? Si A le asigna a su compañero una alta probabilidad de traicionar, le conviene traicionar, pues esto le brinda una menor condena promedio. Por el contrario, si cree que B cooperará con alta probabilidad, entonces, cooperar le proporciona una menor condena promedio.

La paradoja de los dos sobres

El cálculo de promedios puede resultar engañoso, como lo refleja la siguiente paradoja conocida como la ***paradoja de los dos sobres***:

> *Hay dos sobres con dinero sobre una mesa. Ambos sobres están sellados. Se sabe que un sobre contiene el doble de dinero que el otro.*
>
> *Una persona elige uno de los dos sobres y mira su contenido. Luego, se le ofrece la opción de cambiar su elección original y quedarse con el dinero del otro sobre.*
>
> *¿Es conveniente cambiar?*

El razonamiento es el siguiente. La persona elige un sobre. Digamos que contiene 20 doblones. Con probabilidad ½, es el sobre con menos dinero y, por lo tanto, el otro sobre contiene 40 doblones. Por otro lado, con probabilidad ½, es el sobre que contiene más dinero y, por lo tanto, el otro sobre contiene solo 10 doblones.

En el primer caso, al cambiar, la persona gana 20 doblones. En el segundo caso, pierde 10 doblones al cambiar. Luego, con probabilidad ½, la ganancia es de 20 doblones y, con probabilidad ½, la pérdida es de 10 doblones. Por lo tanto, la ganancia promedio al cambiar de sobre es igual a $20 \times ½ - 10 \times ½ = 5$ doblones. Luego, conviene cambiar. Este resultado es completamente absurdo.

Otro hecho curioso: el otro sobre contiene 40 doblones con probabilidad ½ y 10 doblones con probabilidad ½. Luego, en promedio, contiene 25 doblones, es decir, en promedio, contiene más dinero que el sobre elegido inicialmente. Otra vez, esta conclusión es absurda.

El razonamiento correcto es como sigue. Digamos que hay 60 doblones en total: 20 en un sobre y 40 en el otro. Con probabilidad ½, se elige inicialmente el sobre con 20 doblones y se ganan 20 doblones al cambiar. Por otro lado, con probabilidad ½, se elige inicialmente el sobre con 40 doblones y se pierden 20 doblones al cambiar. Luego, la ganancia promedio es 0.

En conclusión, cambiar de sobre o no es, en promedio, irrelevante.

> **Un cálculo más general**: otra vez, el resultado del cálculo anterior no depende del valor exacto de dinero en los sobres. Digamos que el dinero total es igual a 3**X**: **X** en uno de los sobres y 2**X** en el otro. Con probabilidad ½, se elige inicialmente el sobre con **X** y se gana **X** al cambiar. Por el contrario, con probabilidad ½, se elige el sobre con 2**X** y se pierde **X** al cambiar. La ganancia esperada es $\mathbf{X} \times \frac{1}{2} - \mathbf{X} \times \frac{1}{2} = 0$.

La distribución de una variable aleatoria

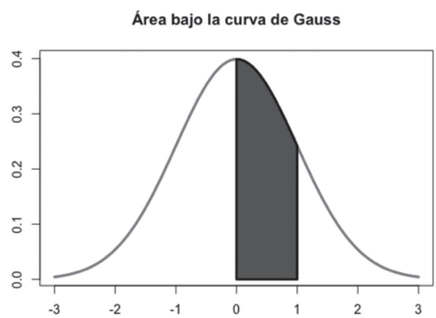
El cálculo de la esperanza de una variable aleatoria consiste en listar todos sus posibles valores y ponderar estos valores por sus respectivas probabilidades. El conjunto de los valores de estas probabilidades se conoce como la **distribución** de la variable aleatoria. Como en el caso de las distribuciones de probabilidad, los términos *distribución* y *ley* en este contexto son sinónimos.

Si una variable aleatoria asume el valor 1 con probabilidad **p** y el valor 0 con probabilidad 1 – **p**, entonces, de alguna manera, le estamos asignando una probabilidad **p** al valor 1 y 1 – **p** al valor 0. Es razonable, por lo tanto, llamar a esta variable aleatoria una variable aleatoria de Bernoulli, al igual que la distribución de Bernoulli. Decimos que se trata de una variable aleatoria de Bernoulli de parámetro **p**. Se puede verificar fácilmente que su esperanza es igual a **p**.

Recordemos que la distribución binomial corresponde al número de experimentos exitosos cuando el experimento se repite **N** veces y tiene probabilidad **p** de ser exitoso. Si codificamos los experimentos exitosos por el valor 1 y los experimentos fallidos por el valor 0, el éxito de un experimento se convierte en una variable aleatoria Bernoulli de parámetro **p**. La suma de estas variables aleatorias es exactamente igual al número de experimentos exitosos. La suma de variables aleatorias es, por supuesto, una variable aleatoria. La distribución binomial corresponde entonces a la distribución de una suma de variables aleatorias Bernoulli y es por lo tanto la distribución de una variable aleatoria que llamaremos variable aleatoria binomial de parámetros **N** y **p**.

Las distribuciones de probabilidad y las variables aleatorias son dos caras de una misma moneda. La distribución de una variable aleatoria es una ley de probabilidad sobre el conjunto de sus posibles valores y tiene un gráfico. El gráfico continuo de la ley gaussiana corresponde asimismo a la distribución de una variable aleatoria continua: la variable aleatoria gaussiana.

Podemos ahora interpretar las probabilidades asociadas a la ley gaussiana. La probabilidad de que una variable aleatoria gaussiana tome algún valor en un intervalo dado, por ejemplo, entre 0 y 1, es igual al área bajo la campana de Gauss limitada por los puntos 0 y 1, tal como se ilustra en la figura.



La esperanza se comporta bien respecto de la suma. La esperanza de la suma de variables aleatorias es igual a la suma de sus esperanzas individuales. Por ejemplo, recordemos que una variable aleatoria binomial de parámetros **N** y **p** es la suma de **N** variables aleatorias Bernoulli de parámetro **p**. La esperanza de esta última es igual a **p**, por lo tanto, la esperanza de la binomial es igual a **N** × **p**. Concretamente, la esperanza de una binomial de parámetros **N** = 20 y **p** = ½ es igual a 20 × ½ = 10. Notemos que esto coincide más o menos con la cima de la campana de su gráfico.

El teorema del límite central y la ley gaussiana

Estudiamos la ley gaussiana en relación con los histogramas de distribuciones de probabilidad discreta. Tal como las distribuciones de probabilidad discreta se pueden entender como la distribución de alguna variable aleatoria, la ley gaussiana corresponde a una variable aleatoria que puede asumir cualquier valor real: una variable aleatoria de ley gaussiana.

Esta ley aparece naturalmente en el estudio de las sumas de variables aleatorias. Ya mencionamos que el promedio aritmético de una gran cantidad de variables aleatorias de igual distribución se acerca a la esperanza. Esta es la ley de los números grandes. El teorema del límite central agrega un grado de precisión.

Recordemos que la ley binomial representa la cantidad de éxitos en un experimento representado una cierta cantidad de veces. Como cada éxito o fracaso se representa por una variable aleatoria de Bernoulli, una binomial se puede entender como una suma de variables de Bernoulli independientes. Así, el teorema de De Moivre asegura que la ley de una suma de variables aleatorias independientes es aproximadamente una variable aleatoria de ley gaussiana. Este es el **teorema del límite central**. Esta afirmación es parcialmente errónea en dos sentidos que dilucidaremos de inmediato.

Cuando graficamos distribuciones binomiales, observamos que la punta de la campana se desplaza hacia la derecha cuando aumenta el valor del parámetro **p**. La campana de Gauss, por otro lado, es perfectamente simétrica. Esto significa que, para obtener la ley gaussiana, tenemos que "centrar" la suma de variables aleatorias, lo que equivale a sustraer su esperanza, que coincide simplemente con la suma de las esperanzas.

Estas consideraciones sencillas dan lugar a una eterna discusión entre matemáticos. ¿Debe enunciarse el teorema como teorema del límite central o teorema central del límite? Los abogados de la primera terminología argumentan a menudo que, dado que se sustrae el promedio, el límite es "centrado". La segunda terminología sugiere que lo central es el teorema; sin duda, es un teorema central en la teoría. La discusión se puede zanjar recurriendo a la historia. El nombre del teorema proviene de un artículo de George Pólya de 1920 intitulado "Sobre el teorema límite central del cálculo probabilístico y del problema de los momentos". Lo central es, por lo tanto, el "teorema límite' o "teorema del límite", un tipo particular de teorema que se refiere a un límite. En este caso, el límite aludido es la ley gaussiana. El término utilizado para nombrar "teorema límite" en alemán es *Grenzwertsatz*, formado a partir de las palabras *Satz* = 'teorema' y *Grenz* = 'frontera' o 'límite'. La terminología más correcta parece ser, por lo tanto, teorema del límite central, entendiéndose que no se refiere a un "límite centrado", sino que a un "teorema del límite" que es central en la teoría de probabilidades.

Ya resolvimos el problema de la posición de la campana. Ahora, cuando graficamos histogramas de distribuciones binomiales, también observamos que estos se vuelven "pequeños" cuando aumenta el parámetro **N**. Este fenómeno debe ser corregido para obtener algo que se parezca efectivamente a una ley gaussiana. Esto se obtiene dividiendo la suma de variables aleatorias por una cantidad adecuada que corresponde a la raíz cuadrada de la cantidad de variables aleatorias sumadas. Combinando ambos mecanismos, concluimos que para obtener el teorema del límite central debemos sustraer la esperanza de la suma de variables aleatorias y dividir el resultado por la raíz cuadrada del número de términos.

El teorema del límite central no se restringe a sumas de variables aleatorias Bernoulli. El matemático Laplace obtuvo una versión más general para sumas de variables aleatorias bastante arbitrarias. Esto se conoce como teorema de De Moivre-Laplace. Dado que Laplace obtuvo la forma más o menos definitiva de este teorema, es sorprendente que la ley gaussiana no se conozca como ley laplaciana. Gauss observó esta ley en relación con los errores de medición en lo que se conoce como ley de los errores. Cuando Laplace se enteró de estos trabajos, pudo relacionar su teorema con la ley de los errores, pero, indudablemente, el descubrimiento de Gauss fue anterior.

La ley de los errores nos da una manera de interpretar el teorema del límite central y la emergencia de la ley gaussiana. Si en una situación la incertidumbre proviene de un gran número de pequeñas contribuciones más o menos independientes, entonces lo observado sigue una ley gaussiana. Aquí, las pequeñas contribuciones corresponden a las variables aleatorias incluidas en la suma.

Esto se puede contraponer a la ley de Poisson, que aparece como el resultado de un gran número de intentos con ínfimas probabilidades de éxito. En el teorema del límite central, cada contribución es pequeña, pero no tan pequeña como en el caso de Poisson. Existen otros teoremas de límite en situaciones en las que la contribución de algunos términos de la suma es considerablemente mayor que las demás, pero esa es una historia completamente diferente relacionada con las distribuciones de colas pesadas.

PROCESOS ESTOCÁSTICOS

Hasta ahora, hemos hablado mayormente de variables aleatorias que describen una situación en un momento determinado. Sin embargo, las probabilidades se pueden utilizar también para describir movimientos influenciados por el azar. Los objetos matemáticos que describen los movimientos aleatorios se conocen como **procesos estocásticos**, siendo el movimiento browniano el máximo exponente de esta familia.

Los procesos estocásticos no describen solo el movimiento. En general, se pueden usar para describir la evolución de un sistema. Se aplican, por ejemplo, para describir el crecimiento de poblaciones o de colonias de bacterias y la variación del precio de las acciones, entre otras cosas.

En los procesos estocásticos, el estado del proceso en cada instante es una variable aleatoria. En este sentido, un proceso estocástico es una colección de variables aleatorias. En general, estas variables aleatorias no son independientes. Si estamos estudiando el presupuesto de un jugador de cara o sello que apuesta cada vez un doblón, sabemos que, jugada tras jugada, su cantidad total de dinero solo puede variar en una unidad. Para describir un proceso estocástico, tenemos que caracterizar su mecanismo de transición, es decir, cómo cambia de manera aleatoria con el paso del tiempo. Este mecanismo está dado por las **probabilidades de transición** y puede ser extremadamente complicado. Estudiaremos con cierto detalle el mecanismo de transición más amigable: los **procesos de Markov**.

La marcha aleatoria

Para introducir nuestro primer ejemplo, volvemos al juego de cara o sello, pero, esta vez, el jugador se moverá a la derecha o a la izquierda dependiendo del resultado de la jugada. Cada vez que obtiene cara, el jugador se mueve un metro hacia la derecha. Por el contrario, cada vez que obtiene sello, se mueve un metro hacia la izquierda. Esto describe el mecanismo de transición de un proceso estocástico conocido como **marcha aleatoria** y, pese a que lo definimos de manera bastante artificial, tiene un rango de aplicación increíblemente amplio.

El ejemplo más sencillo en el que aparece la marcha aleatoria es el presupuesto de un jugador que gana o pierde un doblón dependiendo del resultado de un juego de cara o sello o del color de la ruleta. En un ámbito más complejo, la marcha aleatoria se puede utilizar para modelar de manera sencilla el precio de una acción en la bolsa.

En el mundo de la física, la marcha aleatoria se utiliza para describir la forma de un polímero o para aproximar el movimiento de partículas diminutas suspendidas en un líquido, la trayectoria real siendo representada por el movimiento browniano, que estudiaremos más adelante. En estos casos, la marcha se mueve en tres dimensiones. El mecanismo de transición es muy similar al caso unidimensional que acabamos de describir. En otro contexto, la marcha aleatoria es utilizada por algunas redes sociales para sugerir cuentas a seguir.

La marcha aleatoria es una suma de variables aleatorias independientes, de hecho, variables aleatorias de tipo Bernoulli que asumen los valores 1 y –1 con probabilidad ½. Por lo tanto, de acuerdo con el teorema del límite central, su posición en cada instante sigue aproximadamente la distribución gaussiana. En una dimensión, la marcha aleatoria oscila vigorosamente entre la izquierda y la derecha, y en grandes intervalos de tiempo se mueve arbitrariamente lejos en ambas direcciones. Sin embargo, como los pasos a la derecha se equilibran globalmente con los pasos hacia la izquierda, este movimiento es bastante lento. En 100 pasos, la marcha se desplaza a grandes rasgos unos 10 metros hacia ambos lados. En 10.000 pasos, unos 100. Esto, de algún modo, nos recuerda la ley de los errores y la aparición de la famosa raíz cuadrada.

La marcha aleatoria en dos dimensiones se puede imaginar como el recorrido de un peatón en una ciudad formada por calles perpendiculares que, en cada esquina, decide tomar una de las cuatro direcciones posibles con probabilidad ¼ cada una. En tres dimensiones, tenemos que incluir dos direcciones más: hacia arriba y hacia abajo.

El matemático japonés Kiyoshi Itô describió la marcha aleatoria en dos dimensiones como el recorrido de un hombre borracho que, en cada esquina, decide aleatoriamente hacia dónde ir. Por otro lado, la marcha aleatoria en tres dimensiones describe el recorrido de un pájaro borracho. Debido a las grandes diferencias entre las marchas aleatorias en dimensiones dos y tres, Itô decía, con bastante humor, que un hombre borracho llegará en algún momento a su casa, mientras que un pájaro borracho probablemente nunca encontrará su nido. Esta dicotomía se enuncia en el lenguaje de los procesos estocásticos como *recurrencia en oposición a transiencia*.

La marcha aleatoria también recibe el nombre de *paseo al azar*, de ahí el nombre de este libro.

Cadenas de Markov

Una característica importante de la marcha aleatoria es que, en cada instante, decide aleatoriamente hacia dónde ir, sin importar su trayectoria pasada. Basta con saber dónde se encuentra en el presente para intentar adivinar su posición en el futuro.

El registro de las posiciones pasadas de un proceso estocástico se conoce como la **historia del proceso**. En general, el paso que tomará un proceso estocástico en un instante dado puede depender de toda su historia, pero, en el caso de la marcha aleatoria, su trayectoria futura solo depende de su posición actual. Los procesos cuyo futuro depende solo de su estado presente y no de su historia se conocen como **procesos de Markov**, nombrados en honor al matemático ruso Andreï Markov, un pionero de los procesos estocásticos. En términos coloquiales, se dice que los procesos de Markov **no tienen memoria**.

Veamos algunos ejemplos. Iniciamos nuestra conversación con la marcha aleatoria, que es el paradigma de un proceso de Markov. Por lo tanto, cualquier situación modelada por una marcha aleatoria, como el presupuesto de un jugador o el precio de una acción, constituye un proceso de Markov.

El tamaño de una población de bacterias es, a grandes rasgos, un proceso de Markov. Si tenemos un cultivo en un vaso de Petri, para saber cuántos individuos habrá dentro de una hora, basta con conocer el número actual y la tasa de reproducción y muerte. Este proceso estocástico es bastante interesante, dado que puede resultar tanto en un crecimiento descontrolado de la colonia como en su extinción. El presupuesto de un jugador comparte esta característica: cuando el jugador se queda sin dinero, simplemente deja de jugar, a menos de empeñar su reloj. Estos son ejemplos de procesos de Markov **absorbentes**: existe un estado "absorbente" en el cual se permanece para siempre después de alcanzarlo: la extinción o la bancarrota.

Otra familia de procesos de Markov está dada por los procesos de llegada, como las llegadas sucesivas de clientes en una cola o las horas de llegada de un autobús en un paradero. Estas últimas no siempre son "markovianas". Supongamos que el servicio de buses es extremadamente regular y los buses arriban uno tras otro a intervalos rigurosos de 15 minutos. Al llegar al paradero, solo sabemos que tendremos que esperar a lo más 15 minutos para iniciar nuestro viaje. Sin embargo, si conocemos la hora en que llegó el último bus —una información perteneciente a la historia del proceso— podremos calcular con exactitud la hora de llegada del bus siguiente y, de hecho, de todos los buses siguientes. Por lo tanto, esto no es un proceso de Markov. Por el contrario, si el tiempo que separa la llegada de dos buses consecutivos es aleatorio, el proceso resultante es markoviano en algunos casos. La teoría de colas considera no solo los tiempos de llegada de clientes en una cola, sino que también el tiempo que demoran en ser atendidos. Esto da frecuentemente lugar a procesos de Markov. La teoría de colas tiene numerosas aplicaciones en ingeniería y telecomunicaciones.

El proceso resultante de barajar un mazo de manera aleatoria es un proceso de Markov. Basta con conocer el ordenamiento actual del mazo para describir su distribución tras barajarlo. Pero ¿cuántas veces es necesario barajar un mazo para asegurar que esté bien mezclado? Esta pregunta fue contestada por el mago y matemático Persi Diaconis: basta con barajar siete veces.

Un proceso de Markov se describe a través de sus **probabilidades de transición**. Con esto, podemos formular un modelo markoviano extremadamente sencillo de predicción del tiempo. Basta con especificar las probabilidades de lluvia al día siguiente condicionadas al tiempo de hoy. Por ejemplo, basado en observaciones repetidas, podríamos concluir que la probabilidad de que llueva mañana dado que hoy no llovió es igual a 10% y que la probabilidad de que llueva mañana dado que hoy llovió es igual a 60%. Las probabilidades para un día sin lluvia son iguales a 90% y 40% tras un día con y sin lluvia respectivamente. Esto es un proceso de Markov que, por cierto, no es muy fiable.

Pareciera, no obstante, que los procesos de Markov no son una regla general. La mayoría de los mecanismos de transición no son markovianos. Daremos dos ejemplos sencillos.

En el siguiente ejemplo, suponemos que tenemos en un cofre cuatro monedas de 5 doblones, cuatro monedas de 10 y cuatro monedas de 20 doblones; las sacamos aleatoriamente una tras otra y las disponemos sobre la mesa. El proceso que consideraremos es la cantidad total de dinero sobre la mesa, sin considerar el detalle de las monedas sorteadas. Supongamos que, tras cinco sorteos, tenemos 40 doblones. Con esta información, podemos determinar la distribución del dinero tras el siguiente sorteo. Para formar 40 doblones con cinco monedas, hay solo dos posibilidades: dos de 5, una de 10 y una de 20, y cuatro de 5 más una de 20. En el primer caso, tras el sexto sorteo, podemos tener 45, 50 o 70 doblones. Sin embargo, en el segundo caso, ya agotamos las monedas de 5 doblones y, por lo tanto, al instante siguiente, tendremos al menos 50 doblones sobre la mesa.

El blackjack es un juego de azar en el que es posible establecer una estrategia no markoviana a favor del jugador. Los orígenes del juego son remotos e inciertos. Una forma primitiva del juego se menciona en las *Novelas ejemplares* de Cervantes escritas entre los siglos XVI y XVII. Las reglas del blackjack son complejas, pero, a grandes rasgos, diremos que opone uno o varios jugadores a un crupier y se juega con hasta ocho mazos de 52 cartas simultáneamente. El objetivo consiste en obtener una suma de cartas mayor a la del crupier sin que esta exceda los 21. Un jugador con una suma de cartas mayor a 21 pierde inmediatamente. Las cartas normales valen lo que su número indica, mientras que las sotas, reinas y reyes cuentan como 10, y el as puede valer 1 u 11 según sea conveniente. Inicialmente, se reparten dos cartas a cada jugador. Si un jugador obtiene una suma de cartas igual a 21, gana inmediatamente. El juego se detiene cuando el crupier llega a los 17 o más, en cuyo caso gana quien lo supere. Si el crupier supera los 21, ganan los jugadores. Una vez que termina la jugada, se descartan las cartas utilizadas y se prosigue con las cartas que quedaron en el mazo o los múltiples mazos.

En cada turno, el jugador puede elegir varias alternativas, entre las que se encuentra aceptar una carta adicional. Esta jugada tiene una ventaja y un peligro: la carta adicional puede permitir superar la suma de cartas del crupier, pero arriesga al jugador a sobrepasar los 21, lo que lo elimina de inmediato. El crupier debe sistemáticamente agregar una carta a su mano.

En este juego, es conveniente llevar un registro de las cartas utilizadas. Por ejemplo, si se han utilizados muchas cartas elevadas, podemos pedir cartas adicionales con bajo riesgo de superar los 21. Si, por otro lado, quedan muchos 10 y ases, podemos aumentar nuestra apuesta adicional en cada jugada al existir una alta probabilidad de obtener directamente un 21. Al recordar todas las cartas utilizadas, recurrimos a toda la historia del proceso y decidimos de nuestra estrategia de forma no markoviana.

Existen varias mnemotecnias para recordar lo esencial de las cartas jugadas. Estas técnicas no son ilegales —siempre que no se utilice un dispositivo para aplicarlas—, pero los casinos tienen el derecho de despedir respetuosamente a un jugador sospechado de "contar" cartas con demasiada eficiencia.

Persi Diaconis es un matemático y mago, lo que lo convierte en *matemago*. Trabajó como mago profesional en su juventud y usaba a menudo las matemáticas en el diseño de sus trucos. Volvió a la universidad motivado por sus ganas de aprender probabilidades. Es autor de numerosos trabajos en teoría de números, probabilidades y estadísticas, y coautor de un libro de *matemagia*. Es especialmente conocido por sus artículos acerca de barajar mazos en los que utiliza herramientas de álgebra abstracta y sus conocimientos de trucos de cartas. Fue muy cercano a Martin Gardner.

Serpientes y escaleras

El juego de serpientes y escaleras consiste en un tablero cuadriculado adornado de escaleras y de serpientes. En cada jugada, los jugadores lanzan un dado para determinar cuántas casillas avanzarán. El primero en llegar al otro extremo del tablero gana. La particularidad del juego radica en el efecto de las escaleras y serpientes: cada vez que un peón aterriza en la casilla ocupada por la base de una escalera, se desplaza inmediatamente hasta su otro extremo, adelantando así una cantidad considerable de casillas. Las serpientes, por el contrario, resultan en retrocesos.

Este juego, bajo el nombre de *serpientes y flechas*, se originó en India y llegó a Inglaterra durante el siglo XIX. En sus inicios, tenía una dimensión moral. Las flechas representaban un camino entre un buen comportamiento y su recompensa, y las serpientes, el camino entre un mal comportamiento y sus consecuencias negativas. Así, el estudio lleva al conocimiento y la falta de cuidado, al daño.

En un artículo de 1993, Althoen, King y Schilling analizan este juego usando cadenas de Markov. En efecto, la posición futura de un peón solo depende de su posición actual y del resultado de lanzar el dado. Esto es otro ejemplo de proceso de Markov absorbente donde la última casilla marca el final del juego. Consideraron un tablero con 100 posiciones, con 10 serpientes y 9 escaleras. Concluyen que un jugador demora, en promedio, 39,2 jugadas en cruzar el tablero.

Intuitivamente, se pensaría que agregar escaleras acorta el juego, mientras que agregar serpientes lo alarga. Esto no siempre es el caso. Los autores calcularon que, al agregar una escalera entre las casillas 46 y 94, efectivamente disminuye la duración promedio del juego a 29,8 movimientos. Sin embargo, agregar una escalera entre las casillas 79 y 81 alarga el juego en más de dos jugadas, ya que dificulta el acceso a la escalera que vincula las casillas número 80 y 100. Agregar una serpiente entre las casillas 83 y 7 alarga el juego en alrededor de seis jugadas, mientras que agregar una serpiente entre las casillas 29 y 27 acorta el largo promedio en un poco más de una jugada, al aumentar las chances de alcanzar la escalera que lleva desde la casilla 28 a la 84.

El árbol de Galton-Watson

En la Inglaterra del siglo XIX existía una inquietud por la desaparición de los apellidos ligados a las familias aristocráticas. Estas consideraciones llamaron la atención del polímata sir Francis Galton (1822-1911), quien, junto al reverendo William Watson, creó un modelo matemático para la genealogía descrito en un artículo de 1874: el proceso de Galton-Watson.

El modelo es muy fácil de formular. Existe inicialmente un individuo. Este individuo original da luz a un número aleatorio de "hijos" que sigue alguna distribución de probabilidad. Estos hijos constituyen la primera generación. Cada miembro de la primera generación produce a su vez un número aleatorio de hijos según la misma distribución, independientemente unos de otros. Y el proceso se repite indefinidamente.

El número de individuos en cada generación es, por lo tanto, una suma de variables aleatorias. La cantidad de términos en esta suma es a su vez un número aleatorio; corresponde al número de individuos de la generación anterior, una cantidad aleatoria.

Bajo ciertas condiciones, este proceso lleva a la extinción. Si existe una probabilidad positiva de que los individuos no tengan descendencia, este proceso de "ramificación" puede detenerse. En el contexto original del problema, esto conlleva la desaparición del apellido.

El proceso de Galton-Watson puede verse como un modelo elemental para la evolución de una población de bacterias e incluso para los árboles filogenéticos que trazan la evolución de las especies y explicitan cómo una especie puede ramificarse en varias especies distintas a lo largo del tiempo. Su tratamiento matemático es ameno, bastante explícito y constituye un ejemplo imprescindible en un primer curso matemático de probabilidades a nivel universitario. Se puede representar gráficamente por un árbol, motivo por el cual se lo conoce frecuentemente como *árbol de Galton-Watson*.

Podemos agregar aún más probabilidades a este modelo: podemos imaginar que cada individuo se desplaza aleatoriamente y, en algún momento aleatorio, se asienta y ramifica de acuerdo con las reglas del proceso, después de lo cual sus hijos siguen sus propias marchas aleatorias. Este proceso estocástico —la marcha aleatoria con ramificación— combina tres mecanismos aleatorios: la marcha aleatoria que describe la posición de los individuos, la ramificación aleatoria que determina el número de hijos y el instante de ramificación, que también es aleatorio.

Sir Francis Galton (1822-1911) fue un intelectual inglés multifacético cuyo conocimiento se extendía tanto a la estadística como a la geografía y las ciencias sociales. En probabilidades, se lo conoce principalmente por el proceso de Galton-Watson. Además, introdujo la noción de desviación estándar que cuantifica cuánto se "dispersa" una distribución de probabilidad o un conjunto de datos.

MARCHAS ALEATORIAS EN LA WEB

La matemática es una herramienta omnipresente. No siempre la utilizamos directamente, pero dependemos en cada momento de una gran cantidad de procesos que hacen uso de ella. Al comprar en un supermercado, no calculamos nosotros mismos el valor de la cuenta final, pero el cálculo tiene que ser efectuado por alguna entidad. Al fin y al cabo, esto no es más que una simple operación de aritmética. La situación es dramáticamente más compleja en cualquier interacción que involucre las telecomunicaciones. Volviendo al ejemplo del supermercado, al momento de pagar con tarjeta, se desencadena una sucesión de procesos que permiten comunicar con nuestro banco y efectuar el pago en una fracción de segundo. La información se codifica mediante algoritmos determinados y viaja de manera inmaterial de acuerdo con otra serie de algoritmos. Todo el proceso es guiado por operaciones matemáticas, desde la codificación inicial de la información, su transmisión, su lectura y procesamiento hasta los algoritmos que garantizan la seguridad y confidencialidad de la transacción. Todos estos procesos invisibles descansan últimamente sobre sólidos teoremas e implementaciones robustas que garantizan su buen funcionamiento.

Internet es un espacio imprescindible en nuestra vida cuya complejidad no deja de crecer. La matemática juega un papel preponderante en su funcionamiento y, por supuesto, las probabilidades tienen un rol protagónico. El objeto central en los algoritmos de búsqueda y de sugerencia de contenidos es la marcha aleatoria.

Motores de búsqueda y PageRank

Casi toda búsqueda de información en la web comienza en un motor de búsqueda o buscador. Hoy en día, la cotidianidad de estas herramientas nos impide apreciar su virtud casi mágica. Con solo indicar un par de palabras, el buscador arroja una gran cantidad de sitios cuyo contenido es sorprendentemente acertado.

Quienes recordamos el estado de la web cuando esta recién se estaba popularizando podemos dar fe de la evolución de los buscadores durante las últimas décadas; en un principio, abundaban las sugerencias de escasa calidad. También podemos recordar que varios de ellos competían más o menos en igualdad de condiciones hasta que un monstruo acaparó un protagonismo sin precedente. Es innegable que Google cambió la forma en la que nos relacionamos con la web y transformó la navegación en una experiencia mucho más eficiente.

Los motores de búsqueda son mucho más que buscadores de palabras y es fácil convencerse de que de ninguna manera proceden solamente por conteo de palabras. Por ejemplo, en una búsqueda de la palabra *probability*, el primer sitio arrojado —un sitio pedagógico— exhibe esta palabra 24 veces, mientras que en la segunda entrada —Wikipedia— esta palabra se emplea 138 veces. De hecho, una búsqueda de la palabra *probabilidad* arroja incluso un sitio en el que esta palabra no se utiliza: la entrada de Wikipedia correspondiente a *probability*. La búsqueda de contenidos descansa sobre algoritmos matemáticos que incluyen la probabilidad de manera central y el protagonista de esta revolución es la marcha aleatoria.

La idea fundamental que propició la evolución de los motores de búsqueda fue aprovechar la estructura íntima de la web, es decir, la estructura que nace de los hipervínculos que relacionan las páginas web entre ellas. En este paradigma, el comportamiento de un usuario se puede modelar como una marcha aleatoria o un proceso de Markov que elige estos hipervínculos de manera aleatoria en búsqueda de contenido de su interés: un **navegante aleatorio**.

La diferencia con las cadenas de Markov que descubrimos anteriormente radica en el punto de partida de la navegación. Este es a su vez más o menos aleatorio. La tarea del motor de búsqueda es proveer al navegante un sitio inicial que, con alta probabilidad, concuerde con su búsqueda o lo lleve rápidamente a un sitio de alto interés tras unos pocos clics. El buscador debe, por lo tanto, establecer una medida de probabilidad sobre la web: este es el **PageRank**, calculado usando el algoritmo del mismo nombre.

En palabras de Google, PageRank se basa en la naturaleza democrática de la web utilizando la vasta estructura de vínculos como indicador de la importancia individual de un sitio. Google interpreta un vínculo desde una página A a una página B como un "voto" de A a favor de B. Un sitio con un gran número de votos es, así, un sitio importante. Pero Google no se limita a un mero conteo de estos votos, sino que considera que votos provenientes de páginas importantes tienen una importancia mayor y, a su vez, le transmiten una mayor importancia a la página a la que apuntan. La importancia de un sitio se traduce en el valor de su PageRank. A mayor PageRank, mayor importancia y, por lo tanto, mayor relevancia en una búsqueda.

Esta definición es un tanto circular: para determinar la importancia de un voto, se debe conocer el PageRank del sitio votante, que se calcula en base a la importancia de los votos que recibe. Parece ser un cálculo imposible. La idea para resolver este problema radica en que, al recorrer la web, el navegante aleatorio establece de manera natural una medida de probabilidad sobre ella, como la frecuencia de visitas a un sitio dado durante una larga sesión de navegación. Sitios con muchos vínculos apuntando hacia ellos serán visitados, en promedio, un gran número de veces. Sitios con pocos vínculos podrán ser visitados a menudo siempre y cuando provengan de sitios con alta tasa de visita. El PageRank es, después de todo, una medida de probabilidad y es natural que se pueda calcular utilizando probabilidades.

Google: este nombre proviene de la palabra *googol*, la unidad utilizada para designar el número gigantesco 10^{100}, así como *mil* se refiere a 10^3 y *millón*, a 10^6. Esta compañía fue fundada a fines del siglo pasado por Sergey Brin y Larry Page mientras eran estudiantes de la Universidad de Stanford. La tecnología y los servicios desarrollados por Google van mucho más allá del motor de búsqueda original e incluso trascienden el contexto de internet.

Antes de describir el algoritmo de cálculo de PageRank, tenemos que describir con más cuidado los objetos matemáticos que se utilizarán como modelo de la web: los **grafos**. También discutiremos el modelo matemático del navegante aleatorio: la **marcha aleatoria sobre un grafo**.

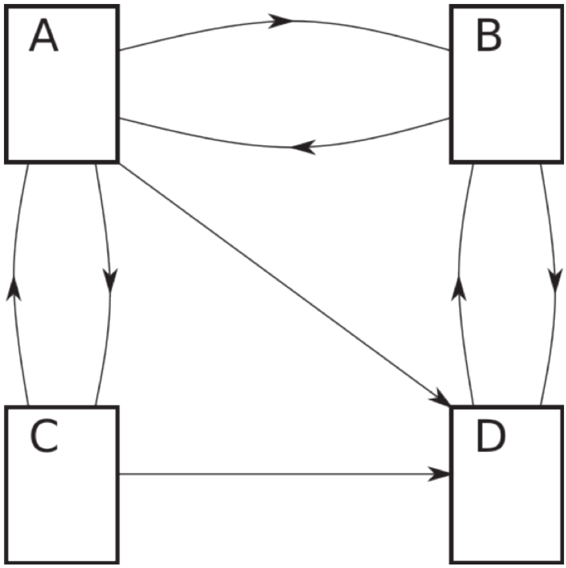
Grafos

Un grafo es un objeto matemático compuesto de dos tipos de entidades relacionadas: los vértices y las aristas. Los vértices son, de alguna manera, los lugares descritos por el grafo. Las aristas son las conexiones o vínculos entre estos vértices: dos vértices están conectados si hay una arista que los une. En este caso, decimos que los vértices son vecinos.

Un ejemplo cotidiano de grafo está dado por la red de metro de una ciudad. Los vértices corresponden a las estaciones y las aristas, a los tramos de las líneas. Así, dos estaciones son vecinas —existe una arista que las une— si son estaciones sucesivas a lo largo de una misma línea. Algunas estaciones solo tienen dos vecinos: la estación anterior y la estación siguiente. Otra, aquellas que se encuentran a la intersección de varias líneas, tienen una mayor cantidad de vecinos. Al recorrer la red de metro, los viajeros se desplazan de estación en estación vecina, describiendo un camino a lo largo del grafo.

La web es otro ejemplo de grafo. Los vértices corresponden a las páginas, y las aristas, a lo hipervínculos. Este ejemplo es, sin embargo, distinto al ejemplo del metro. En el metro, una arista conecta las estaciones A y B en ambos sentidos. En la web, los hipervínculos apuntan en una única dirección. Así, un sitio A puede tener un vínculo hacia un sitio B sin que el sitio B apunte a su vez hacia el sitio A. En la web, las aristas o vínculos pueden ser vistas como flechas que apuntan desde la página de partida hasta la página de llegada. Los grafos en los que las aristas tienen una dirección se conocen como **grafos dirigidos**.

El ejemplo siguiente muestra una red de cuatro páginas que utilizaremos como una simplificación un tanto aberrante de internet. Las cuatro páginas A, B, C y D, están representadas por rectángulos y los ocho hipervínculos, por flechas. Por ejemplo, la página A tiene tres hipervínculos, apuntando a las tres otras páginas. En el otro extremo, la página D tiene un único hipervínculo apuntando hacia B, no obstante, es apuntada por todas las otras páginas. Nos referiremos a este ejemplo como la web de cuatro páginas.

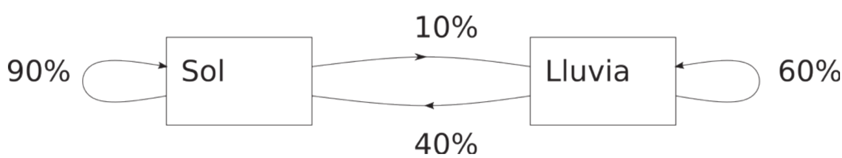


Los siete puentes de Könisberg y la teoría de grafos: el origen histórico de la teoría de grafos se puede encontrar en el problema de *los siete puentes de Könisberg* y su resolución por el matemático Leonhard Euler. La ciudad de Könisberg contaba con dos islas unidas al resto de la ciudad mediante siete puentes. El problema consistía en encontrar un camino a través de la ciudad que cruzara cada uno de los puentes exactamente una vez. Euler demostró que esto es imposible. En su demostración, Euler sentó los cimientos para la teoría de grafos y una disciplina matemática conocida como topología que hace abstracción de la forma de los objetos, sintetizando solo sus propiedades más relevantes. El matemático Paul Erdös es recordado, entre otras cosas, por sus numerosas contribuciones a la teoría de grafos.

La marcha aleatoria sobre un grafo

En lo que sigue, consideraremos solamente grafos dirigidos, es decir, grafos cuyas aristas apuntan en una dirección determinada.

En un capítulo anterior, consideramos un modelo sencillo de predicción del tiempo. Recordemos que supusimos que la probabilidad de que un día soleado sea sucedido por otro día soleado es de 90%, mientras que la probabilidad de que un día de lluvia suceda a un día de lluvia es de 60%. La probabilidad de las otras figuras —lluvia después de sol y sol después de lluvia— se calcula por paso al complemento. Este sistema se representa por un grafo con dos vértices: Sol y Lluvia. Las cuatro flechas representan las cuatro posibles transiciones, sus respectivas probabilidades siendo indicadas cerca de las flechas. Así, si iniciamos la marcha en "Sol", hay una probabilidad de 90% de que, tras el primer salto, la marcha siga en el sitio "Sol", es decir, se desplaza por la flecha a la izquierda del esquema. Por el contrario, con probabilidad 10%, la marcha saltará al sitio "Lluvia".



Para nuestro modelo de la web, consideraremos dos simplificaciones. Primero, no consideraremos flechas que apuntan desde una página en sí misma. Segundo, supondremos que la probabilidad de seguir cualquiera de las aristas que sale de un vértice es uniforme. Las probabilidades de transición de una marcha aleatoria sobre este grafo se calculan, por lo tanto, de la manera siguiente: si la marcha se encuentra sobre el sitio X, contamos el número de sitios que reciben una flecha que sale de X y la marcha salta a cualquiera de ellos con la misma probabilidad.

Así, en la web de cuatro páginas, la página B tiene dos hipervínculos: hacia A y D. Luego, si el navegante aleatorio se encuentra en la página B, saltará con probabilidad ½ a la página A y con probabilidad ½ a la página D. Notemos otra vez que la página D tiene un único hipervínculo apuntando hacia B. Si el navegante se encuentra en D, saltará hacia B con probabilidad 1.

Para modelar el navegante aleatorio, es necesario especificar el punto de partida de la marcha de manera aleatoria. Especificamos, por lo tanto, una medida de probabilidad sobre los vértices del grafo y disponemos la marcha de acuerdo con estas probabilidades. La manera más sencilla consiste en escoger la medida de probabilidad uniforme sobre los vértices. En la web de cuatro páginas, supondremos que la marcha se inicia en cualquiera de las cuatro páginas con probabilidad ¼.

Si iniciamos la marcha aleatoria según la probabilidad uniforme sobre los vértices y esperamos el primer salto de la marcha, podemos calcular la probabilidad de que la marcha se encuentre en cualquiera de los vértices tras este salto. En la web de cuatro páginas, ¿cuál es la probabilidad de que el navegante se encuentre en la página B tras el primer salto? Para que esto ocurra, el navegante debe encontrarse inicialmente en las páginas A o D, ambas posibilidades ocurriendo con probabilidad ¼. Si se encuentra inicialmente en A, salta a B con probabilidad ⅓. Luego, la probabilidad de encontrarse inicialmente en A y saltar a B es igual a ¼ × ⅓ = 1/12. Si se encuentra inicialmente en D —lo que, recordemos, ocurre con probabilidad ¼—, la probabilidad de saltar a B es igual a 1, dado que el único hipervínculo naciente de D apunta hacia B. Luego, la probabilidad de encontrarse inicialmente en D y saltar a B es igual a ¼ × 1 = ¼. Las dos probabilidades recién calculadas suman 1/12 + ¼ = ⅓. Vemos que la probabilidad de encontrarse en B tras el primer clic es mayor que la probabilidad de encontrarse inicialmente en esa página. Esto se debe a la estructura de hipervínculos de nuestra red.

Podemos repetir este ejercicio y calcular la probabilidad de que la marcha se encuentre sobre cualquier vértice dado tras una cantidad determinada de saltos.

El procedimiento anterior nos entrega una secuencia de medidas de probabilidad sobre el grafo correspondiente a las posiciones sucesivas de la marcha. En principio, estas medidas de probabilidad son todas distintas; de hecho, lo hemos verificado en un ejemplo sencillo. Existe, no obstante, una medida de probabilidad privilegiada conocida como la **medida de equilibrio** que satisface la propiedad siguiente: si iniciamos la marcha con estas probabilidades, entonces su posición tras el primer salto —y, de hecho, tras cualquier número de saltos— se distribuye otra vez según la medida de equilibrio. En la web de cuatro páginas, estas probabilidades se pueden calcular fácilmente. No detallaremos el cálculo, que solo involucra operaciones aritméticas elementales, y nos limitaremos a dar el resultado. Las probabilidades de encontrarse inicialmente en A, B, C y D con la medida de equilibrio son iguales a 3/13, 5/13, 1/13 y 4/13 respectivamente.

La medida de equilibrio codifica una información importante respecto de la marcha sobre el grafo. Bajo esta medida, la probabilidad de un vértice corresponde a la frecuencia de visitas a este vértice durante una larga caminata. En el ejemplo, vemos que, de acuerdo con la medida de equilibrio, B recibe una gran probabilidad. Eso es bastante natural: tras cualquier visita a D, el navegante debe necesariamente saltar a B. Es decir, B es un paso obligado. Por otro lado, la página C recibe una probabilidad bajísima. Otra vez, esto es bastante natural: la única posibilidad de llegar a C es desde A y, aun así, un navegante en A solamente salta a C con probabilidad ⅓. Notemos también que, pese a que la página D es accesible desde todas las otras páginas, su probabilidad es menor que la de B.

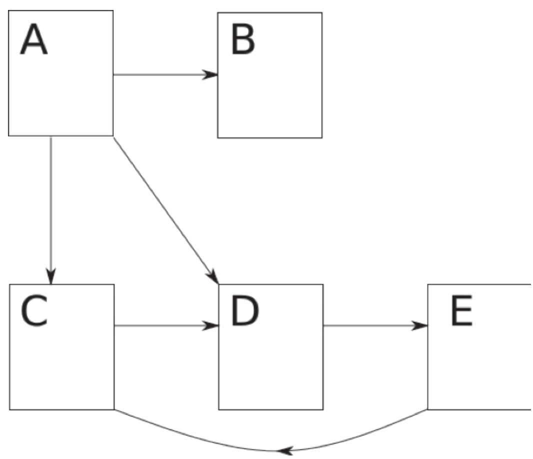
Calcular la medida de equilibrio de una marcha puede ser complejo. Conceptualmente, se trata de resolver un sistema de ecuaciones bastante sencillo, pero la cantidad de ecuaciones crece con el tamaño del grafo, volviendo este cálculo muy trabajoso. Afortunadamente, existe una forma sencilla de aproximar esta medida de equilibrio. Así como la medida de equilibrio refleja la frecuencia de visitas a los vértices del grafo, también aparece como la distribución de la posición de la marcha tras un gran número de saltos. En cierto sentido, la distribución de la posición de la marcha *converge* a la medida de equilibrio. En una gran cantidad de situaciones, esta convergencia es además bastante rápida.

Luego, para lograr una aproximación razonable de la medida de equilibrio, basta con dejar evolucionar la marcha una cantidad razonable de saltos y observar la distribución de probabilidad de su posición. En la práctica, podemos proceder como sigue: iniciamos 100 marchas aleatorias independientemente según la medida uniforme sobre los vértices y las dejamos saltar 20 veces. Luego, graficamos la medida de probabilidad empírica de la posición de las marchas como aprendimos en un capítulo anterior. Salvo situaciones patológicas, esto nos entrega una muy buena aproximación de la medida de equilibrio.

El algoritmo PageRank

Disponemos de todas las herramientas para definir apropiadamente el PageRank, que, recordemos, cuantifica la importancia de una página web. Consideramos el grafo cuyos vértices son todas las páginas web y cuyas aristas son todos los hipervínculos. Luego, consideramos la marcha aleatoria sobre el grafo resultante. Esta marcha aleatoria tiene una medida de equilibrio que es una medida de probabilidad sobre las páginas web. A grandes rasgos, esto es el PageRank. Hay, sin embargo, un par de complicaciones que tenemos que resolver para formular nuestro modelo definitivo para el navegante aleatorio.

Hay al menos dos situaciones que podrían alterar nuestro cálculo de PageRank. La estructura de vínculos podría generar **ciclos**, es decir, el navegante podría estar forzado a recorrer una y otra vez una misma sucesión de páginas, tal como lo muestra el ejemplo siguiente. Vemos que, si el navegante llega a las páginas C o D, quedará eternamente navegando entre C, D y E.



Algunas páginas podrían también actuar como trampas. Es decir, podrían no contener ningún vínculo. Así, una vez que el navegante llega a ellas, no puede seguir navegando. Este es el caso de la página B en el ejemplo anterior.

La solución propuesta por Brin y Page para solucionar estos problemas se condice con el comportamiento de un navegante aleatorio más realista. Un navegante no se limita a seguir una secuencia de vínculos. En algún momento, puede decidir que la búsqueda actual es infructuosa y devolverse al buscador. Es decir, se reinicia la marcha aleatoria. En consecuencia, nuestro modelo tiene que integrar un parámetro adicional: la probabilidad de reiniciar la búsqueda desde cero. Podemos, por lo tanto, formular nuestro modelo definitivo para el navegante aleatorio: el navegante escoge su punto de partida siguiendo cierta distribución de probabilidad determinada. Tras cada salto, decide proseguir con la marcha aleatoria con cierta probabilidad. Con la probabilidad complementaria, decide reiniciar la búsqueda y escoge un nuevo punto de partida según la ley de probabilidad con la que inició la marcha. Esto describe una nueva marcha aleatoria que tiene una medida de equilibrio. Esto es precisamente el PageRank.

Para calcular el PageRank, posicionamos inicialmente el navegante aleatorio con la distribución uniforme sobre las páginas y lo dejamos evolucionar siguiendo las reglas anteriores. Tras una cantidad sorprendentemente pequeña de saltos, obtendremos una buena aproximación de la medida de equilibrio. En su artículo original, Brin y Page estiman que en una web de 322 millones de vínculos basta con observar al navegante tras 52 saltos. En situaciones prácticas, la probabilidad de reiniciar la marcha se escoge cercana al 15%.

El método de Monte-Carlo: el cálculo de la medida de equilibrio involucra una gran cantidad de operaciones aritméticas sencillas que pueden ser realizadas por un computador. Para un grafo de gran tamaño, el tiempo requerido para efectuar este cálculo puede ser gigantesco, incluso con un computador de alta capacidad. El problema del tiempo de cálculo es transversal a un gran número de situaciones. A menudo, es posible diseñar un proceso estocástico que nos provee una solución aproximada del cálculo deseado de manera rápida y bastante precisa a través de simulaciones. Esto se conoce como el método de Monte-Carlo y se refiere a un conjunto de técnicas que permiten utilizar las probabilidades para resolver problemas que incluso no involucran el azar en su formulación original. El método fue ideado en los años cuarenta por el matemático y físico polaco Stanislaw Ulam (1909-1984) mientras trabajaba en el Proyecto Manhattan. El nombre se refiere a un famoso casino de Mónaco frecuentado por su tío.

Marchas aleatorias en redes sociales

Las redes sociales sentaron un nuevo paradigma en internet. El desafío mayor consiste en mantener y ampliar la base de usuarios activos. Esto se logra en parte ayudando a usuarios nuevos y antiguos a descubrir contenidos interesantes. El éxito de las redes sociales se basa, por lo tanto, en un acertado sistema de sugerencia de contenidos. Tal como en el caso de los motores de búsqueda, los algoritmos de sugerencia de contenidos tienen una alta componente probabilística.

La definición de contenido interesante comparte con el PageRank el estar basada en una noción algo recursiva y autorreferente. Un contenido se puede considerar interesante para un usuario dado si es interesante para usuarios similares a él. A su vez, dos usuarios se pueden considerar como similares si comparten intereses semejantes. Esto se refleja en que usuarios similares tienden a seguir a los mismos usuarios o a usuarios que son similares a su vez.

En su esencia, la sugerencia de contenido debe ser personalizada, es decir, se debe basar en una suerte de PageRank individualizado. Esto se logra encontrando un modelo adecuado de navegante aleatorio a través de una red social. Este navegante comparte muchas de las características del navegante aleatorio encontrado anteriormente. Hay, sin embargo, una diferencia fundamental que radica en la forma en la que el usuario de la red social reinicia su búsqueda. Siempre vuelve a empezar en el mismo lugar: en su propio espacio.

La marcha aleatoria definida para el cálculo del PageRank debe ser modificada de manera evidente: tras cada salto, el navegante debe volver a su propio espacio con cierta probabilidad y volver a iniciar la marcha. Esto se conoce como marcha aleatoria egocéntrica y su medida de equilibrio es el PageRank personalizado.

Hay otra diferencia entre las redes sociales e internet en cuanto al tipo de sitios que se pueden considerar de interés. En las redes sociales, un sitio es interesante no solo si contiene información directamente relevante, sino también si apunta hacia sitios o usuarios que despliegan dicha información. Existen, por lo tanto, dos tipos de usuarios: los *Authorities,* que proveen información valiosa en su contenido, y los *Hubs*, que apuntan hacia sitios y usuarios interesantes. Tal como la calidad de una página se puede cuantificar con el PageRank, la calidad de un Hub o Authority se puede cuantificar con un debido indicador. Un buen Hub se caracteriza por apuntar a buenos Authorities y un buen Authority es seguido por buenos Hubs.

El algoritmo utilizado para *rankear* los Hubs y Authorities es similar al algoritmo de PageRank e involucra marchas aleatorias. El procedimiento es más complejo, dado que debe arrojar dos indicadores distintos. A grandes rasgos, se requiere considerar dos copias del grafo. En una de ellas, los usuarios serán evaluados según su calidad como Hub y, en la otra, como Authority. Enseguida, se consideran dos marchas aleatorias. La primera se inicia en el mundo de los Authorities, visita los Hubs y se devuelve. Así, esta marcha recorre las dos copias del grafo y tiende a visitar más frecuentemente a Authorities a las cuales apunta una gran cantidad de Hubs de buena calidad. La segunda marcha se inicia en el mundo de los Hubs, visita los Authorities y se devuelve. Este algoritmo se conoce como *Stochastic Approach for Link-Structure Analysis* o SALSA. Los indicadores de Hub y Authority están dados por las medidas de equilibrio de estas marchas aleatorias.

La red Twitter estableció un algoritmo de sugerencia de contenidos que resume todas las ideas anteriores de forma armónica: el algoritmo Who-to-follow. Utiliza PageRank y SALSA, pero se diferencia de este último en su definición más restrictiva de Hubs y Authorities. Su funcionamiento se puede desglosar en varias etapas.

En la primera etapa, se establece el Círculo-de-confianza de un usuario *rankeando* los usuarios que sigue según su relevancia. Esto se logra calculando el PageRank personalizado utilizando la marcha aleatoria egocéntrica. Este Círculo-de-confianza constituye el universo de los Hubs. El universo de los Authorities está conformado por los usuarios seguidos por los miembros del Círculo-de-confianza.

Luego, se utiliza el algoritmo SALSA para evaluar la calidad de los Hubs y Authorities identificados previamente. En esta lógica, un buen Hub es un usuario similar al usuario de partida y un buen Authority se convierte en una sugerencia de contenido.

La teoría de redes: las ciencias, y las matemáticas en particular, están fuertemente guiadas por las necesidades de la época en la que se enmarcan. Así, las civilizaciones agrícolas, ante la necesidad de aprehender adecuadamente las variaciones de luminosidad a lo largo del año, desarrollaron algún tipo de conocimiento geométrico aplicable al estudio del movimiento de los astros. Más recientemente, el desarrollo de la máquina de vapor impulsó nuestro conocimiento de los mecanismos del calor y de su conversión en movimiento: la termodinámica. Cada época tiene sus desafíos específicos y la nuestra no está exenta de ellos. Se observó que una gran cantidad de fenómenos cotidianos provenientes de la biología, la física, la computación e incluso la sociología y la economía exhiben un denominador común: en ellos, una gran cantidad de actores interactúan de manera compleja a través de una intrincada red de relaciones. La teoría de redes es una ciencia interdisciplinaria y con aplicaciones diversas que aporta una visión global de estos fenómenos. Uno de los pioneros de esta disciplina es el físico Albert-László Barabási (1967). Descubrió ciertas propiedades estructurales de internet que se vuelven a encontrar en redes de naturaleza distinta. Sus trabajos tienen aplicación en la biología, la medicina y la sociología.

EL MOVIMIENTO BROWNIANO

En los capítulos anteriores hemos discutido acerca de los procesos estocásticos que permiten, entre otras cosas, modelar los movimientos sujetos al efecto del azar. Nuestro ejemplo central es la marcha aleatoria, cuyo universo de aplicación, ya lo vimos, es extremadamente vasto.

Cerraremos nuestra discusión histórica de las probabilidades con el rey de los procesos estocásticos: el movimiento browniano. Este proceso surgió en la biología, tuvo repercusiones dramáticas en la física, reapareció de forma inesperada en las finanzas, se convirtió en un objeto central de las matemáticas y forma la base de un cuerpo cada vez más amplio de aplicaciones de la teoría de la probabilidad.

Las observaciones de Robert Brown

Robert Brown (1773-1858) fue un botánico escocés y, de hecho, uno de los botánicos más influyentes de su época. Fue un pionero en el uso de microscopios para el estudio riguroso de la biología y, tras un largo viaje en Australia, catalogó miles de especies de plantas desconocidas hasta entonces. El nombre de Brown quedó grabado en un sitio de honor del monumento de la teoría de probabilidades por un artículo publicado en 1828 intitulado "Breve recuento de observaciones microscópicas de partículas contenidas en polen de plantas". Los resultados de Brown permearon mucho más allá del mundo de la biología e iniciaron una cadena de ideas fundamentales en física y matemáticas que, entre otras cosas, llevó a la confirmación definitiva de la naturaleza atómica de la materia. El proceso estocástico más importante lleva su nombre: el **movimiento browniano**, y describe el movimiento aleatorio de los cuerpos microscópicos.

En 1827, Brown realizó una observación sorprendente. Brown estudiaba el proceso de fertilización en flores cuando, al observar partículas de polen suspendidas en un líquido con microscopio, notó que estas describían un movimiento oscilatorio errático y aparentemente impredecible. Su primera hipótesis fue que se trataba de células masculinas dotadas de movilidad propia. Brown, sin embargo, no se detuvo ahí y se propuso verificar —o descartar— su hipótesis inicial mediante una serie de experimentos. Al pulverizar otras partes de la planta, observó invariablemente el mismo fenómeno. Concluye que estas partículas en movimiento corresponden a componentes fundamentales de la vida orgánica que describen un movimiento propio. Otra vez, no se detuvo ahí.

Para descartar que el movimiento observado se originase en partículas vivas, realizó el mismo experimento con plantas petrificadas o conservadas en herbarios por siglos, observando siempre el mismo fenómeno. Rápidamente, sospechó que la materia inorgánica podría tener la misma propiedad, lo que confirmó al repetir el experimento con polvo de metal, rocas e incluso fragmentos de la Esfinge de Guiza.

Era importante determinar si el movimiento era causado por las partículas o por el fluido que las contenía y, en el segundo caso, descartar el efecto de corrientes dentro del fluido. El ingenio de Brown llegó hasta a desarrollar una preparación consistente en una sola partícula de polen contenida en una gota de agua aislada en aceite. Esto descartaba explicaciones mecánicas sencillas como algún tipo de movimiento causado por la evaporación del agua y, de paso, alguna interacción física desconocida entre partículas distintas.

La teoría de Brown se puede resumir de esta forma: toda la materia está formada por diminutas partículas —que nombró *moléculas activas*— que describen un movimiento oscilatorio errático e impredecible originado en las mismas partículas.

El procedimiento de Brown es un ejemplo sublime del método científico. Tras sus primeras observaciones, Brown formula una hipótesis que descarta tras realizar nuevos experimentos. La información proveniente de los nuevos experimentos alimenta nuevas hipótesis que pueden ser verificadas o descartadas, hasta llegar a un planteamiento satisfactorio.

Si bien Brown fue el primero en documentar este fenómeno de manera sistemática, el movimiento browniano debe haber sido un dolor de cabeza para cualquiera que intentara observar una suspensión de partículas diminutas. Hay algunos reportes escuetos del movimiento browniano en la literatura de los siglos anteriores, ligados al desarrollo de microscopios en Holanda durante el siglo XVII.

El movimiento browniano apareció muchísimo antes, aunque en contexto y palabras distintas, en un poema de Lucrecio de los años sesenta antes de Cristo. En "De la naturaleza de las cosas", Lucrecio describe la danza de las partículas de polvo iluminadas por un haz de luz. Lucrecio imagina que el aire, como todas las cosas, está constituido de pequeñas partículas en movimiento. Estas partículas imprimen movimiento a partículas más grandes que, a su vez, animan partículas mayores, hasta llegar a las partículas de polvo, sacudidas por este baño de partículas elementales.

Curiosamente, ¡esta es la explicación del movimiento browniano! Contrariamente a la teoría de Brown, el movimiento browniano no se origina en las partículas de polen, sino en el fluido aledaño, él mismo constituido por moléculas que oscilan. Esta conclusión fue alcanzada tras una serie de experimentos realizados por físicos de la segunda mitad del siglo XIX que explicaremos a continuación. El movimiento browniano es una propiedad intrínseca de la materia cuya explicación, casualmente, es muy cercana a la idea poética de Lucrecio.

El movimiento browniano en la física

En los 30 años que siguieron, el trabajo de Brown parece haber suscitado escaso interés y, en particular, no logró llamar la atención de los físicos. No obstante, su artículo parece haber logrado cierta popularidad, incluso fuera de los círculos científicos, como lo sugiere una alusión anecdótica al mismo en la novela *Middlemarch* de la escritora George Eliot publicada en 1872.

El físico y matemático Christian Wiener (1826-1896) realizó una serie de experimentos en 1863 que aportan precisiones a las observaciones de Brown. Wiener logra descartar sistemáticamente explicaciones mecánicas sencillas para el movimiento browniano, como movimiento dentro del fluido debido a evaporación o convección, o interacciones entre las partículas. Contrariamente a Brown, Wiener sugiere que el origen del fenómeno se debe buscar en el fluido y no en las mismas partículas. La teoría de Wiener tiene sin embargo serias fallas, pues, además de invocar la vibración de las moléculas del fluido, alude a la vibración del éter en una longitud de onda correspondiente al color rojo, una explicación difícil de sustentar.

La originalidad de Wiener radica en el intento de explicación molecular del fenómeno. Hoy, todos estamos familiarizados con el hecho de que la materia en todos sus estados está formada por átomos y afirmar que el movimiento browniano se debe al movimiento de estos átomos no resulta difícil de creer. Debemos recordar que, en los tiempos de Wiener, aún se discutía la naturaleza molecular de la materia —es decir, que la materia está constituida de pequeñas unidades— y algunas de las mentes más brillantes de la época favorecían la teoría de que la materia es un continuo. Como veremos más adelante, el estudio del movimiento browniano —en especial, la teoría de Einstein y los experimentos de Perrin— fue crucial en zanjar esta discusión.

Los primeros trabajos experimentales rigurosos sobre el movimiento browniano se deben al físico francés Louis Georges Gouy (1854-1926). Junto con reforzar la hipótesis molecular, estos trabajos permiten extraer las propiedades fundamentales de las trayectorias de las partículas, que se pueden resumir en los siguientes puntos:

1.- El movimiento es extremadamente irregular y las trayectorias no parecen tener tangente.
2.- Las partículas brownianas incluso cercanas describen trayectorias independientes.
3.- El movimiento es más intenso para partículas más pequeñas.
4.- La naturaleza y la densidad de las partículas no tienen relevancia.
5.- El movimiento es más intenso en fluidos menos viscosos.
6.- El movimiento es más intenso a mayores temperaturas.
7.- El movimiento no se detiene.

El primer punto tiene particular relevancia para la matemática del problema. Matemáticamente, se puede reformular diciendo que las trayectorias de las partículas no son diferenciables. Explicar esto con detalle va más allá del alcance de este libro y nos limitaremos a recordar que la diferenciabilidad es una propiedad crucial en la descripción del movimiento de los cuerpos en la mecánica de Newton y que su ausencia es el reflejo de la alta irregularidad de las trayectorias. El fenómeno apela, por lo tanto, a una matemática nueva y, en particular, cualquier teoría matemática acertada del movimiento browniano tiene que incluir esta ausencia de diferenciabilidad. Tendremos que esperar hasta 1920 para ver completado este programa.

El segundo punto implica, entre otras cosas, que el movimiento browniano no se debe a ningún tipo de interacción entre las partículas. Los cuatro puntos siguientes son observaciones cualitativas que se deben poder inferir a partir de cualquier teoría física sólida del fenómeno.

El último punto es un tanto más vago. Después de todo, ¿cómo podríamos afirmar que el movimiento no se detiene basado observaciones durante un intervalo de tiempo finito? Existen algunos reportes anecdóticos de preparaciones conservadas por décadas en las que se pudo constatar la permanencia del fenómeno. El movimiento browniano se observó incluso en gotas de agua atrapadas en cuarzo por milenios.

Saltando casi 2.000 años desde el poema de Lucrecio, llegamos al año 1905, que, en ciencias, se conoce como "*Annus Mirabilis*". En este año, el físico Albert Einstein publicó cuatro artículos que cambiaron la historia de la ciencia al introducir ideas fundamentales en el desarrollo de la física del siglo XX. Los más famosos son, sin duda, el artículo en el que introduce la teoría de la relatividad especial y el artículo en el que aparece la ecuación más famosa de las ciencias: ¡$E = mc^2$! Junto con la teoría de la relatividad general, estos dos trabajos son las contribuciones más célebres de Einstein. El tercer artículo, sobre el efecto fotoeléctrico, es fundamental en nuestra comprensión de la naturaleza de la luz.

En el cuarto artículo, Einstein predice que partículas suspendidas en un líquido debieran describir un movimiento aleatorio debido a las fluctuaciones térmicas del fluido. El artículo de Einstein es la primera teoría del movimiento browniano, la cual se basa en la conciliación de dos puntos de vista aparentemente contradictorios: por un lado, la mecánica de fluidos que considera la materia como un continuo y, por otro lado, la teoría cinética del calor, más familiar en el contexto de los gases. Insistimos sobre el hecho de que la teoría de Einstein es una predicción teórica de un fenómeno observable en principio. En 1905, Einstein no estaba al tanto de los trabajos de Brown y de Gouy.

El trabajo de Einstein consta esencialmente de dos partes. En la primera, Einstein vislumbra la naturaleza gaussiana del fenómeno. El movimiento de una partícula suspendida en un fluido resulta de una gran cantidad de choques con las moléculas del agua, cada uno de los cuales le imprime una ínfima cantidad de movimiento. Así, el movimiento browniano es la suma de una gran cantidad de pequeñas contribuciones más o menos independientes, por lo que cae en el ámbito de la ley de los errores o teorema del límite central. El desplazamiento de una partícula browniana sigue, por lo tanto, una distribución gaussiana.

En la segunda parte, Einstein obtiene una fórmula para la intensidad del movimiento en función de ciertas cantidades elementales. Esta fórmula es muy sencilla y permite explicar fácilmente las observaciones de Gouy. La naturaleza cuantitativa de esta teoría facilita también su verificación experimental.

Poco después de la publicación de su artículo, un colega le dio a conocer los trabajos de Gouy en los cuales Einstein reconoce sus predicciones. No solo las propiedades cualitativas de las trayectorias observadas son coherentes con la teoría, sino que las mediciones se ajustan perfectamente a las fórmulas. En un artículo escrito a finales de 1905, Einstein generaliza su teoría del fenómeno, que ahora nombra *movimiento browniano*. En este trabajo, Einstein también incluye rotaciones aleatorias de las partículas, el movimiento browniano rotacional.

Recordemos que, en esa época, aún se discutía si la materia estaba conformada por pequeñas unidades elementales o si se trataba de un "continuo". La teoría de Einstein aportó evidencias a favor de la naturaleza atómica de la materia y abrió las puertas al diseño de experimentos precisos para ser verificada. Estos experimentos fueron realizados por el físico francés Jean Baptiste Perrin (1870-1942) a inicios del siglo XX. Al verificar la teoría de Einstein, sus experimentos confirman la naturaleza atómica de la materia: la materia está constituida por átomos. Es más, estos átomos vibran, fenómeno que se observa en la oscilación visible de partículas de mayor tamaño: es el movimiento browniano, documentado por Brown y soñado por Lucrecio. Perrin recibió el Premio Nobel de Física por sus trabajos en 1926. Por supuesto, Einstein también recibió el Nobel —en 1921—, aunque su teoría del movimiento browniano, al igual que las partículas de polen, se baña en un mar de creaciones científicas maravillosas que no solo cambiaron nuestra manera de ver el mundo, sino que ampliaron nuestro horizonte hasta límites insospechados.

Notemos que el físico Marian Smoluchowski (1872-1917) llegó a conclusiones similares a las de Einstein en 1906 siguiendo argumentos más probabilísticos. En su trabajo, Smoluchowski reduce el estudio de la trayectoria de la partícula al comportamiento de una marcha aleatoria, un punto de vista crucial en la teoría de probabilidades.

Hasta este punto, la teoría del movimiento browniano se limitaba a formular sus propiedades en base a las observaciones y, en el caso de los trabajos de Einstein y Smoluchowski, a justificar teóricamente el fenómeno partiendo de primeros principios. En matemáticas, esto no es suficiente y se requiere de una construcción rigurosa de los objetos estudiados. Volveremos a este punto, pero antes, romperemos la cronología de este relato para relatar una aparición inesperada del movimiento browniano en un área muy lejana.

El movimiento browniano en la bolsa de comercio

El movimiento browniano fue descubierto independientemente por Louis Bachelier (1870-1946) en sus investigaciones de los mercados financieros. En su tesis de 1900 intitulada "Teoría de la especulación", Bachelier postula que los incrementos en el precio de las acciones siguen aproximadamente una ley gaussiana.

De pequeño, Bachelier trabajó en la bolsa de comercio y asistió posteriormente a las cátedras del matemático Henri Poincaré (1852-1912). En cierto sentido, Bachelier aportó la primera formulación matemática del movimiento browniano. También fue quizás el primero en introducir la marcha aleatoria.

Sus trabajos fueron difícilmente valorados por los matemáticos de la época y su tesis solo recibió una mención honrosa. Esto se debió probablemente a la escasa familiaridad de la comunidad matemática con la economía, y su tesis, aun siendo visionaria, carecía de los elementos rigurosos requeridos por un trabajo puramente matemático. Sin embargo, Poincaré recalcó en varias oportunidades la importancia de estas ideas. La aplicación de la teoría de la probabilidad a los mercados financieros abrió las puertas a un área de investigación extremadamente fructífera: las matemáticas financieras.

Hoy en día, los modelos probabilísticos juegan un papel preponderante en las transacciones bursátiles. Una descripción detallada de estos métodos requeriría de un capítulo completo. Nos limitaremos a mencionar que las matemáticas financieras han sido el motor de desarrollo de una gran cantidad de trabajos teóricos profundos sobre los procesos estocásticos y han sido por décadas un territorio fértil de aplicaciones de la teoría de la probabilidad.

El movimiento browniano en las matemáticas

Hasta acá, el movimiento browniano carecía de una formulación matemática rigurosa, la teoría existente limitándose esencialmente a formular sus propiedades y a justificar su emergencia. La matemática es la ciencia del rigor y no es en general suficiente con enunciar una lista de propiedades para sustentar una teoría. Es absolutamente necesario demostrar que existe un objeto matemático que refleja las propiedades exhibidas por el movimiento browniano. El paso crucial en esta dirección fue completado por el matemático norteamericano Norbert Wiener (1894-1964), también padre de la cibernética y de la inteligencia artificial.

Wiener viajó a Cambridge en 1913 motivado por estudiar lógica con Bertrand Russell, pero este último lo instó a asistir a las cátedras de Godfrey Hardy y a leer los artículos de Einstein. En 1923, Wiener publicó un artículo intitulado "Espacios diferenciales", en el que construye una medida de probabilidad sobre trayectorias bajo la cual los incrementos —o diferencias— siguen una distribución normal. También deduce que estas trayectorias no son diferenciables, justificando matemáticamente las observaciones de Gouy.

De ahí en adelante, el movimiento browniano se convirtió en un objeto medular de las matemáticas. La segunda mitad del siglo XX presenció una explosión de investigaciones en torno a sus propiedades, notablemente de la mano de las escuelas francesa y japonesa. Nacieron las ecuaciones diferenciales estocásticas, que combinan la probabilidad con las ecuaciones diferenciales, la herramienta clásica para la descripción del movimiento de los cuerpos. El movimiento browniano es un objeto de estudio contemporáneo en las matemáticas y sus aplicaciones.

El movimiento browniano es fuente de un sinnúmero de propiedades exóticas. Por ejemplo, si diseccionamos la trayectoria browniana en trozos muy pequeños, veremos que cada trozo observado con lupa es a su vez un movimiento browniano. En este sentido, el movimiento browniano es un fractal y, aún más, un fractal aleatorio.

Las aplicaciones del movimiento browniano son infinitas. En la próxima sección, relataremos el papel de movimiento browniano en la dinámica de los seres microscópicos.

Kiyosi Itô (1915-2008) fue un matemático japonés autor de contribuciones cruciales en la teoría de los procesos estocásticos. Uno de los teoremas más utilizados en el contexto del movimiento browniano lleva su nombre: el *lema de Itô*. Este teorema permite realizar cálculos extraordinariamente limpios para procesos estocásticos y constituye, junto a la integral estocástica, las bases del llamado *cálculo de Itô*. Estas herramientas son fundamentales para el estudio de las ecuaciones diferenciales estocásticas que combinan las ecuaciones de la mecánica clásica con los procesos estocásticos. Las ideas de Itô son aplicadas en una gran variedad de disciplinas, incluyendo la física y las finanzas.

Nadadores microscópicos

Hemos visto que el movimiento browniano describe el movimiento de todo lo diminuto. En sus experimentos, Brown descartó que este fenómeno se debiera a la movilidad de organismos vivos y los físicos tras él concluyeron que su origen se encuentra en las oscilaciones intrínsecas del fluido aledaño. Sin embargo, existen seres diminutos dotados de movilidad propia que son tan pequeños que su trayectoria es dramáticamente afectada por las vibraciones de las moléculas de agua. Se trata de las bacterias y, en realidad, de cualquier organismo microscópico.

Los microbios fueron formalmente descubiertos en el siglo XVII por el científico holandés Antonie van Leeuwenhoek (1632-1723) durante el llamado Siglo de Oro de la ciencia y tecnología holandesa. Su existencia y, en particular, su incidencia en la transmisión de enfermedades se intuía desde mucho antes.

El movimiento del agua y de todo aquello que se desplaza en el agua es el ámbito de la mecánica de fluidos y está gobernado por las ecuaciones de Navier-Stokes. Estas ecuaciones modelan el comportamiento de los fluidos —gases y líquidos—, desde la brisa más suave hasta los torrentes más turbulentos. Como muchas ecuaciones de la física, las ecuaciones de Navier-Stokes se pueden escribir en un par de líneas, pero son increíblemente difíciles de resolver. ¡El Instituto Clay incluso ofrece un premio de un millón de dólares a quien sea capaz de crear una teoría matemática satisfactoria para su estudio!

El mundo visto por una diminuta bacteria es, sin embargo, muy distinto al que vemos nosotros. Un nadador "macroscópico" como un humano o un pez sabe que, si deja repentinamente de nadar, seguirá avanzando por el agua unos instantes, un fenómeno que se conoce como inercia. Este desplazamiento se detiene pronto debido a la viscosidad del fluido que se opone al movimiento. Resulta que la viscosidad, o la manera en que la sentimos, depende de algún modo de nuestro tamaño. Un nadador microscópico sufre los efectos de una viscosidad que percibe esencialmente como infinita. Si deja de nadar, se detiene instantáneamente; no hay inercia. En este contexto, las ecuaciones de Navier-Stokes se simplifican y se conocen simplemente como ecuación de Stokes.

Las bacterias, al ser igualmente o más pequeñas que las partículas de polen de Brown, están sometidas al movimiento browniano. Su trayectoria es una combinación de su propio mecanismo de nado —típicamente, cilios y flagelos— gobernado por la ecuación de Stokes y de las oscilaciones del líquido que producen el movimiento browniano. La teoría de Einstein permite calcular con precisión la intensidad de esta "componente" browniana. Por lo tanto, la trayectoria de un nadador microscópico se debe modelar usando ecuaciones diferenciales estocásticas, que combinan elementos mecánicos y aleatorios.

Nuestro conocimiento del mundo microscópico se vio reforzado con el advenimiento de los microscópicos digitales en la segunda mitad del siglo XX. El procesamiento digital permite, entre otras cosas, estudiar con precisión la trayectoria de las bacterias, individual o colectivamente. Las teorías matemáticas que fusionan la mecánica de fluidos con la probabilidad a través del movimiento browniano retratan las observaciones de manera extraordinariamente acertada.

TODO SE TRATA DE CONTAR: PROBABILIDAD Y COMBINATORIA

Recordemos que un evento es un conjunto de alternativas o, en nuestra terminología matemática, un conjunto de elementos del espacio muestral. Bajo la distribución uniforme, su probabilidad es la razón entre posibilidades favorables y las posibilidades totales. Luego, para el cálculo de probabilidades uniformes, todo radica en ser capaz de contar estas dos cantidades.

Aprendemos a contar desde muy pequeños. Es aparentemente sencillo: listamos los objetos que queremos contar, nos aseguramos de que están todos en la lista, que no sobra nada y simplemente contamos. Para ir más rápido, podemos juntar los objetos de a pares o tríos, pero ¿qué pasa cuando la cantidad de elementos que queremos contar es astronómicamente grande? Por ejemplo, ya dijimos que el número de posibles manos de póker es 2.598.960. En principio, tendríamos que listar todas las posibilidades y contarlas. La lista es gigantesca. ¿Cómo asegurarnos de que no olvidamos alguna alternativa? ¿Cómo estar seguro de que no repetimos ninguna? Claramente, necesitamos un método más rápido y mejor.

La herramienta matemática que nos ayuda a contar en situaciones complejas es la ***combinatoria***. La combinatoria identifica una serie de situaciones típicas que aparecen en el proceso de contar y nos provee de fórmulas o métodos para obtener rápidamente el resultado.

Veremos las nociones básicas de combinatoria: la regla del producto y el cálculo de permutaciones y combinaciones. Algunas de estas nociones ya se utilizaron de manera intuitiva en los capítulos anteriores en situaciones concretas. Entenderlas de mejor manera nos abre las puertas al cálculo de probabilidades en contextos más complicados, como las probabilidades en el juego de póker.

Combinaciones, permutaciones y las leyes elementales de la combinatoria

Esta sección es sin duda la más avanzada desde el punto de vista matemático. Es posible, sin embargo, lograr un entendimiento intuitivo de las principales nociones abordadas aquí —las permutaciones y combinaciones— sin entrar en los detalles. Estas herramientas de cálculo se utilizarán fuertemente en las próximas secciones acerca de la distribución binomial y del póker con dados y con cartas. Otra vez, el lector puede leer la próxima sección sin entrar en los detalles matemáticos.

Empecemos con un ejemplo concreto: en una cafetería, se venden deliciosos bagels. El cliente debe elegir entre cuatro tipos de pan: simple, amapola, sésamo o arándano. Luego, debe elegir el acompañamiento: mermelada, salmón ahumado o pollo teriyaki. Estas alternativas se pueden combinar de todas las maneras posibles. Se puede, por ejemplo, elegir para el desayuno un bagel de arándano con mermelada o, para el almuerzo, un contundente bagel cubierto de semillas de amapola con salmón ahumado. Pero ¿cuántas opciones distintas existen?

Aquí nos encontramos con un principio general de la combinatoria: las posibilidades de escoger dos objetos se multiplican. Esta es la ***regla del producto***. Ya nos encontramos con este carácter multiplicativo cuando discutimos la urna.

Hay cuatro posibilidades para el pan y tres posibilidades para el acompañamiento. Luego, por la regla del producto, hay $4 \times 3 = 12$ recetas distintas, es decir, 12 distintas maneras de combinar pan y acompañamiento. El lector incrédulo podrá hacer el ejercicio de listarlas todas.

Ahora, volvemos a una situación más genérica y muy familiar en probabilidades: veremos cómo aplicar la regla del producto al problema de la urna, con o sin reposición.

Recordemos el contexto. Tenemos una urna con cinco bolas numeradas del 1 al 5. En la primera modalidad, cada vez que elegimos una bola al azar, la volvemos a colocar dentro de la urna. Es el problema de la urna con repetición. Supongamos que escogemos tres bolas secuencialmente: escogemos la primera, la reponemos, escogemos la segunda, y así. ¿De cuántas maneras se pueden escoger las tres bolas? Hay cinco posibilidades para la primera, cinco para la segunda y cinco para la tercera. Por lo tanto, hay $5 \times 5 \times 5 = 5^3 = 125$ posibilidades para la secuencia de tres números. En el lenguaje de las probabilidades, hay 125 posibilidades totales de escoger tres números de este modo y, por lo tanto, la probabilidad de seleccionar cualquier secuencia dada —por ejemplo, 3, 5 y 2— es igual a 1/125 = 0,8%.

Ahora, consideremos la urna sin reposición. En este caso, seleccionamos una bola, la apartamos y seleccionamos la siguiente. Hay cinco posibilidades para la primera bola. Pero, una vez que la seleccionamos, solo quedan cuatro bolas en la urna. Una vez que seleccionamos una de estas, quedan tres bolas en la urna. Luego, las posibilidades totales son $5 \times 4 \times 3 = 60$. La probabilidad de escoger de este modo la secuencia 3, 5 y 2 es, por lo tanto, $1/60 \approx 1{,}7\%$. Estos resultados coinciden con los cálculos del capítulo 3.

Hay un problema con los cálculos anteriores. Nos permiten contar secuencias ordenadas de números. Por ejemplo, la secuencia 3, 5 y 2 y la secuencia 3, 2 y 5 son contadas como dos posibilidades distintas. En muchas circunstancias, es deseable contar estas dos secuencias como una sola, es decir, como una sola ***combinación*** de números. Por ejemplo, en la lotería, si escogimos los números, 6, 9, 13, 23, 25 y 35, poco importa si salieron en ese orden o, digamos, en el orden 13, 23, 9, 35, 25 y 6.

> **La Biblioteca de Babel**: volvemos al cuento de Borges mencionado en el capítulo 3. Asumiendo que un libro consta de 410 páginas de 40 renglones de exactamente 80 caracteres y sabiendo que existen 25 caracteres —22 letras, el espacio, la coma y el punto—, ¿cuántos libros posibles se pueden hallar en la Biblioteca? ¿Cuál es la probabilidad de encontrar un libro en blanco?

Volvamos a la urna sin reposición, que es como una mini lotería con cinco números de los cuales se escogen tres, y digamos que escogemos los números 2, 3 y 5. Estos números pueden aparecer en varios órdenes posibles. El número de órdenes en el que los números 2, 3 y 5 pueden aparecer se conoce como el número de ***permutaciones*** de los números 2, 3 y 5. Las distintas permutaciones son (2, 3, 5); (2, 5, 3); (3, 2, 5); (3, 5, 2); (5, 2, 3) y (5, 3, 2). Hay seis permutaciones de los números 2, 3 y 5. En general, hay seis posibles permutaciones de tres objetos distintos.

En el problema de la urna sin reposición, ¿cuántas combinaciones de tres números existen? Nuestro cálculo anterior nos daba 5 × 4 × 3 = 60 secuencias ordenadas de números en la que cada permutación de tres números es contada seis veces, una vez por cada permutación. Por lo tanto, tenemos que dividir el número encontrado anteriormente por 6. Hay 60/6 = 10 combinaciones de tres objetos distintos escogidos entre cinco objetos distintos.

Supongamos que queremos contar el número de permutaciones de 25 objetos distintos. No parece práctico generar la lista. Este es, de hecho, un número con 26 dígitos. Busquemos una fórmula que nos permita contar rápidamente el número de permutaciones. Resulta que hay un patrón bastante simple.

Si tenemos dos objetos, A y B, hay solo dos permutaciones: primero A y luego B, y lo contrario. Luego, el número de permutaciones de dos objetos es igual a 2. Ya vimos que el número de permutaciones de tres objetos es igual a 6. Es fácil verificar que hay 24 permutaciones de cuatro objetos; basta con listar las distintas permutaciones. La siguiente tabla indica el número de permutaciones de hasta cinco objetos distintos. La tercera columna recalca un patrón bastante curioso.

Número de objetos	Resultado	Cálculo
2 objetos	2	2 × 1
3 objetos	6	3 × 2 × 1
4 objetos	24	4 × 3 × 2 × 1
5 objetos	120	5 × 4 × 3 × 2 × 1

Por analogía, podemos concluir que el número de permutaciones de 10 objetos es 10 × 9 × 7 × 6 × 5 × 4 × 3 × 2 × 1 = 3.628.800. Tenemos, por lo tanto, una manera eficiente de calcular el número de permutaciones de un número arbitrariamente grande de objetos.

El factorial: la operación de multiplicar todos los números desde 1 hasta un número dado se conoce como el factorial. Se denota con un signo de exclamación. Por ejemplo, el factorial de 5 es $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$. El factorial de un número se vuelve extraordinariamente grande a medida que este número crece. Como lo hemos visto, el factorial aparece a menudo en combinatoria.

Este método no es sorprendente y se deduce a partir de una idea que hemos explotado ya más de una vez. En el caso de las permutaciones de 10 objetos, hay 10 posibilidades para el primero, nueve para el segundo, ocho para el tercero y así sucesivamente. Como estas posibilidades se multiplican, obtenemos el resultado anterior.

Volvamos al problema de las combinaciones. Queríamos determinar el número de combinaciones de tres números escogidos entre el 1 y el 5, es decir, tres números distintos sin importar el orden. Vimos que hay $5 \times 4 \times 3 = 60$ secuencias ordenadas de tres números que tenemos que dividir por el número de permutaciones de estos tres números. El número de dichas permutaciones es $3 \times 2 \times 1 = 6$. Concluimos que hay $60/6 = 10$ combinaciones de tres números escogidos entre el 1 y el 5.

Apliquemos estas ideas a la lotería: tenemos que escoger seis números entre el 1 y el 41 y no nos importa el orden en el que aparecen al momento del sorteo. Calculemos. El número de secuencias ordenadas de seis números escogidos entre el 1 y el 41 es:

$$41 \times 40 \times 39 \times 38 \times 37 \times 36 = 3.237.399.360$$

El número de permutaciones de seis números distintos es:

$$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

En nuestro cálculo del número de secuencias, cada combinación de seis números se contó, por lo tanto, 720 veces. Para obtener el número de combinaciones, se debe dividir el número de secuencias por el número de permutaciones de seis números. Así, el número de combinaciones de seis números escogidos entre el 1 y el 41 es: $3.237.399.360/720 = 4.496.388$.

Paul Erdös (1913-1996), un matemático húngaro, fue uno de los científicos más prolíficos del siglo XX, autor de más de mil artículos en conjunto con un número gigantesco de colaboradores . Su trabajo se centró especialmente en la matemática discreta, en la que la combinatoria juega un rol preponderante. Erdös fascina tanto por su extensa producción científica como por su modo de vida peculiar. Llevaba una vida errante cuyo único fin aparente era la matemática. Viajaba de conferencia en conferencia y solía visitar a sus colegas alrededor del mundo por unos días, el tiempo suficiente para producir algunos artículos. Luego, seguía su camino. En matemáticas, se habla del número de Erdös de una persona. Este mide los grados de separación con Erdös. Una persona que colaboró directamente con él tiene un número de Erdös igual a uno. Una persona que colaboró con un colaborador de Erdös tiene un número igual a 2 y así sucesivamente. Erdös apreciaba mucho las demostraciones bellas y elegantes. Cuando se encontraba con una de ellas, decía que era una demostración *del libro*, aludiendo a un supuesto libro en el cual Dios consigna las mejores demostraciones de teoremas.

La distribución binomial, otra vez

Disponemos de todas las herramientas para calcular explícitamente las probabilidades asociadas a la distribución binomial. Para simplificar la discusión, fijaremos los parámetros **N** = 10 y **p** = 0,7. En este caso, la distribución binomial representa el número de experimentos exitosos cuando el experimento es repetido 10 veces de manera independiente y cuando cada experimento tiene una probabilidad de éxito igual a 0,7.

Calculemos la probabilidad de obtener exactamente cuatro experimentos exitosos. Primero, calculemos la probabilidad de que los experimentos número 2, 5, 8 y 9 sean exitosos y todos los demás resulten en fracaso. Los experimentos exitosos contribuyen con una probabilidad igual a $0{,}7 \times 0{,}7 \times 0{,}7 \times 0{,}7 = (0{,}7)^4$, número que se debe multiplicar por la probabilidad de que los otros seis experimentos sean infructuosos: $0{,}3 \times 0{,}3 \times 0{,}3 \times 0{,}3 \times 0{,}3 \times 0{,}3 = (0{,}3)^6$. La probabilidad de que los experimentos señalados sean exitosos y los demás fracasen es entonces igual a $(0,7)^4 \times (0,3)^6 \approx 0,02\%$. Notemos que hemos usado fuertemente la independencia de los distintos experimentos.

Ahora, tenemos que calcular de cuántas maneras se pueden elegir los experimentos que serán exitosos. Tenemos 10 posibilidades para elegir el primero de ellos, nueve para el segundo, ocho para el tercero y siete para el cuarto. Hay entonces $10 \times 9 \times 8 \times 7 = 5.040$ maneras de escoger cuatro experimentos entre 10. Ahora, con este procedimiento, estamos contando de más. Como discutimos anteriormente, las secuencias (2, 5, 8, 9) y (2, 9, 8, 5) son contadas como dos entidades distintas, siendo que representan el mismo grupo de cuatro experimentos. Tenemos que dividir por el número de permutaciones de cuatro objetos, que es igual a $4 \times 3 \times 2 \times 1 = 24$. El número de posibles combinaciones de cuatro experimentos exitosos entre 10 experimentos totales es, por lo tanto, igual a $5.040/24 = 210$.

Volviendo a la probabilidad buscada, la probabilidad de obtener exactamente cuatro experimentos exitosos entre 10 con probabilidad de éxito igual a 0,7 es igual a $210 \times (0,7)^4 \times (0,3)^6 \approx 0,04$.

Con el mismo procedimiento, concluimos que la probabilidad de obtener exactamente siete experimentos exitosos es aproximadamente igual a 0,27, un valor bastante mayor al anterior. Notemos que 7 corresponde a la esperanza de una variable aleatoria binomial de parámetros $\mathbf{N} = 10$ y $\mathbf{p} = 0,7$ y coincide más o menos con la posición de la cima del gráfico de su distribución.

Existe una manera más eficiente de calcular estas probabilidades empleando otro objeto central de la combinatoria: los coeficientes binomiales.

Precalentamiento: el póker con dados

Como su nombre lo indica, el póker de dados es una variante del póker que se juega con dados, más precisamente con cinco dados cuyas caras corresponden al 9, 10, la sota, la reina, el rey y el as. Ya repasaremos las denominaciones de la baraja inglesa. Los distintos resultados obtenidos al lanzar los cinco dados simultáneamente siguen un orden de importancia similar a las manos de póker que consideraremos en la próxima sección.

El mejor resultado posible consiste en obtener cinco figuras idénticas, por ejemplo, que los cinco dados arrojen reyes. ¿Cuál es la probabilidad de obtener esta configuración?

Contemos primero todas las configuraciones posibles. Hay seis posibilidades para cada dado. Por la regla del producto, el número de posibilidades para cinco dados debe ser igual a 6 × 6 × 6 × 6 × 6 × 6 = 6^5 = 7.776.

Hay solo seis maneras de obtener el mismo resultado en todos los dados, que corresponden a las seis figuras impresas en las caras del dado. La probabilidad de obtener esta configuración es, por lo tanto, igual a 6/7.776 ≈ 0,07%.

La siguiente configuración en orden de importancia es aquella en la que exactamente cuatro de los dados arrojan el mismo resultado, por ejemplo, As-As-As-As-9. Para elegir la figura que se repetirá cuatro veces, tenemos seis posibilidades y solo cinco para la segunda figura. Por lo tanto, hay 6 × 5 = 30 configuraciones distintas de dos figuras distintas, una de las cuales se repetirá cuatro veces. Ahora, tenemos que escoger cuáles de los cuatro dados arrojarán la figura repetida. Para esto, basta con contar el número de combinaciones de cuatro objetos escogidos entre cinco. Para elegir cuatro dados uno tras otro, tenemos 5 × 4 × 3 × 2 = 120 posibilidades. Enseguida, tenemos que dividir por el número de permutaciones de estos cuatro dados, que es igual a 4 × 3 × 2 = 24. El número de combinaciones de cuatro objetos escogidos entre cinco es entonces igual a 120/24 = 5. Combinando todo nuestro razonamiento, hay 30 × 5 = 150 distintas configuraciones de cinco dados con exactamente cuatro figuras idénticas. La probabilidad de obtener una de estas es, por lo tanto, igual a 150/7.776 ≈ 1,93%, donde recordamos que 7.776 corresponde al número total de configuraciones posibles para los cinco dados.

Después de este precalentamiento, podemos iniciar nuestros cálculos de probabilidades para las manos de póker.

Póker: cálculos para valientes

Recordemos lo básico: un mazo de cartas consiste en 52 cartas distintas. Hay cuatro colores distintos: espadas, corazones, rombos y tréboles. Hay 13 cartas de cada color. Estas cartas están numeradas del 1 al 12. Algunos números tienen una denominación especial. El 1 se conoce como as, el 11 como sota, el 12 como reina y el 13 como rey. Una mano de póker consiste en cinco cartas distintas. En abstracto, una mano de póker es un sorteo en una urna con 52 objetos de los cuales se escogen cinco sin reposición. Los cálculos para el póker son muy similares al caso de los dados, pero involucran números muchísimo más grandes. Contemos el número total de manos de póker.

El número de secuencias de cinco cartas escogidas una tras otra es:

$$52 \times 51 \times 50 \times 49 \times 48 = 311.875.200$$

El número de permutaciones de cinco cartas distintas es $5 \times 4 \times 3 \times 2 \times 1 = 120$. En el número de más arriba, cada combinación de cartas se contó, por ende, 120 veces. Luego, para obtener el número de manos de póker, es decir, el número de combinaciones de cinco cartas, se debe dividir este número por 120. Por lo tanto, el número de manos de póker es

$$311.875.200/120 = 2.598.960$$

Las manos de póker tienen una jerarquía. El jugador que obtiene la mano de mayor valor gana. La peor mano es aquella en la que ningún tipo de carta se repite y no se da ningún patrón particular, como cartas ordenadas del 2 al 6, que se conoce como *escalera*. La probabilidad de obtener una mano de este tipo es aproximadamente 50,11%. Es decir, hay una probabilidad un poco mayor a ½ de no obtener nada. Esta probabilidad no es tan fácil de calcular, por lo que omitimos estos detalles por ahora. La siguiente mano en importancia es el par, aquella en la que se obtiene exactamente un par; por ejemplo, el 3 de espadas, el 3 de diamantes y, digamos, un 2 de cualquier color, una sota de cualquier color, y un 6 de cualquier color. La probabilidad de obtener un par es aproximadamente 42%. Ya daremos indicios de cómo calcular esta probabilidad más abajo. La probabilidad restante es ocupada por las manos de mayor valor. Las probabilidades de obtener alguna de ellas son tan pequeñas que usaremos porcentajes y, aun así, obtendremos números con muchos decimales.

La mejor mano de póker es la escalera real. Consiste en cuatro cartas consecutivas del 10 al rey, del mismo color, y el as del mismo color. Hay solo cuatro de estas. La probabilidad de obtener una de ellas es muy baja. Es igual a 4 dividido por el número total de manos de póker: aproximadamente 0,000154%.

La siguiente mano en importancia es la escalera de color. Consiste en cinco cartas consecutivas del mismo color que no formen una escalera real. Notemos que la secuencia sota, reina, rey, as y 2 no constituye una escalera. Por así decirlo, la escalera no puede "dar la vuelta", excepto, de algún modo, la escalera real. Para contar las escaleras de color tenemos que tomar en cuenta las posibilidades de color —son cuatro— y los posibles puntos de partida, que son las cartas del 1 al 9; el 10 como carta inicial da lugar a una escalera real. Hay, por lo tanto, 4 × 9 = 36 escaleras de color. La probabilidad de obtener una de ellas es aproximadamente 0,00139%.

En orden decreciente de valor, nos encontramos con el póker: cuatro cartas de un mismo número. Hay 13 posibilidades para escoger el número —del as al rey— y 52 – 4 = 48 posibilidades para escoger la carta restante. Luego, hay 13 × 48 = 624 pókeres distintos. La probabilidad de un póker es aproximadamente 0,024%.

La siguiente mano es el *full house*: un trío y un par. Primero, tenemos que escoger los dos números que formarán respectivamente el par y el trío y, luego, contar el número de posibles *full houses* formados con esos números.

Para elegir el número que formará el par, hay 13 posibilidades. Quedan 12 posibilidades para escoger el número que formará el trio. Esto nos da 13 × 12 = 156 posibilidades.

Una vez que escogimos estos números, tenemos que contar la cantidad de manos posibles. Digamos que queremos contar el número de *full houses* formados por un par de 7 y un trío de 5. Hay cuatro cartas con el número 7. Necesitamos el número de combinaciones de dos de estas cartas. Ya aprendimos a calcular este número. Hay 4 × 3 = 12 maneras en las que pueden salir dos 7 sucesivamente. Hay que dividir este número por el número de permutaciones de dos objetos que son 2. Hay, por lo tanto, 12/2 = 6 maneras de obtener un par de 7. Ahora, para escoger un trío de 5, necesitamos contar el número de combinaciones de tres objetos escogidos entre 4. Hay 4 × 3 × 2 = 24 maneras en las que pueden aparecer los tres 5 sucesivamente. El número de permutaciones de tres objetos es 6. Luego, hay 24/6 = 4 maneras de escoger un trío de 5. Juntando estos números, hay seis maneras de obtener un par de 7 y cuatro maneras de obtener un trío de 5. Hay, entonces, 6 × 4 = 24 maneras de obtener un par de 7 y un trío de 5.

El 7 y el 5 son arbitrarios. Vimos que hay 156 maneras de elegir estos dos números. El número total de *full houses* es 156 × 24 = 3.744. La probabilidad de *full house* es 0,14%.

Terminamos esta sección calculando la probabilidad de la siguiente mano en orden de importancia: la escalera. Esto es cualquier combinación de cinco números sucesivos que no dan lugar a una escalera de color y, por lo tanto, a una escalera real si el número de inicio es el 10.

Hay 10 posibilidades de escoger el número inicial. Digamos que escogimos el 2. Hay cuatro posibilidades de escoger un 2, cuatro posibilidades de escoger un 3 y así hasta llegar al 6. Hay, por lo tanto, $4 \times 4 \times 4 \times 4 \times 4 = 4^5 = 1.024$ escaleras que parten del 2. De estas, cuatro son escaleras de color que tenemos que descartar. Hay, entonces, 1.020 escaleras que parten de un número del 1 al 10 y 10 posibilidades para escoger este número de partida. Hay en total $1.020 \times 10 = 10.200$ escaleras. La probabilidad de una escalera 0,39%.

Observamos que la mejor manera de calcular la probabilidad de obtener la peor mano consiste en sumar las probabilidades de obtener manos de mayor valor y pasar al complemento.

Probabilidad de par: dejamos un ejercicio para valientes. ¿Cuál es la probabilidad de obtener un par y nada más? Primero tenemos que escoger el número que formará el par. Luego, contar el número de pares formados con este número. Luego, escoger tres números distintos al primero y distintos entre sí para las cartas restantes y, finalmente, contar de cuántas maneras puedo elegir cartas con cada uno de estos números. El número de manos de este tipo es 1.098.240. Dejamos los detalles del cálculo para el lector.

Para finalizar, proponemos cuatro problemas clásicos de probabilidad. El primero tiene una solución muy compleja que puede ser formulada de manera elemental, pero cuya demostración requiere de herramientas matemáticas complicadas. Nos limitaremos a formular la solución dando algunas pistas de su validez e invitando al lector a estudiarla en casos más sencillos. El segundo problema está al alcance de un estudiante de cálculo con nociones elementales de probabilidad, sin embargo, no detallaremos el cálculo que lleva a la solución. El tercer problema tiene una solución que puede ser verificada cuidadosamente caso por caso. El último problema es un problema de ingenio.

El problema de los 100 prisioneros

El siguiente problema fue formulado en 2003 por Peter Bro Miltersen y se conoce como el *problema de los 100 prisioneros*.

> *En una cárcel, hay 100 prisioneros, numerados del 1 al 100, que esperan ser ejecutados. El alcaide decide darles una última oportunidad.*
>
> *En una sala, dispone un mueble con 100 cajones también numerados del 1 al 100. Distribuye 100 papeletas numeradas del 1 al 100 en los cajones según algún orden arbitrario.*
>
> *Cada prisionero entra a la sala y tiene derecho a abrir 50 cajones en cualquier orden. Después de esto, se cierran los cajones e ingresa el siguiente prisionero.*
>
> *Si cada prisionero encuentra su propio número en uno de los 50 cajones que abrió, serán todos perdonados. En caso contrario, serán ejecutados según estaba planificado.*

Los prisioneros tienen derecho a discutir una estrategia antes de iniciar el proceso, pero no se pueden comunicar durante este.

¿Cuál es la mejor estrategia que seguir?

La probabilidad de que un prisionero encuentre su propio número al abrir 50 cajones al azar es igual a ½. Si los prisioneros no planifican una estrategia, estos eventos son todos independientes y la probabilidad de que todos los prisioneros encuentren su propio número es igual a ½ multiplicado con sí mismo 100 veces. Esto es igual a un cero, una coma, luego 30 ceros y, finalmente, un dígito distinto de cero. Esta estrategia resultará sin duda en un fracaso.

Existe una estrategia que permite asegurar una probabilidad de sobrevivencia de cerca del 30% y consiste en lo siguiente: cada prisionero abre, en primer lugar, el cajón numerado con su propio número. Si su número se encuentra en ese cajón, ya no hace falta seguir. Si el cajón contiene otro número, abre el cajón correspondiente a ese número y así sucesivamente.

Si bien no trabajaremos los detalles de esta solución, la observación fundamental es que la disposición de los números en los cajones es una ***permutación*** de los números del 1 al 100. Podemos pensar que, originalmente, cada número estaba en el cajón con su propio número y que, en algún momento, el contenido de los cajones fue permutado al azar. Esto es, por lo tanto, una permutación aleatoria. Todas las permutaciones posibles de los números del 1 al 100 forman, con derecho propio, un espacio muestral gigantesco sobre el cual se puede definir la probabilidad uniforme.

Para barajar los números, una forma sensata de proceder es la siguiente. El alcaide saca el número 1 de su cajón y elige un destino al azar, digamos, el 57. Pone el 1 en el cajón 57 y elige el destino del número 57 y así sucesivamente.

Las permutaciones tienen ***ciclos***. Al permutar los números, el 1 es cambiado por otro número que, a su vez, es cambiado por otro, y así sucesivamente. En algún momento, el alcaide deberá elegir el cajón número 1 para guardar algún número de la secuencia y se cierra el ciclo, y se inicia un ciclo nuevo. La cantidad de pasos necesarios para cerrar el ciclo es el largo del ciclo.

Al seguir la estrategia propuesta, los prisioneros estarán recorriendo los ciclos de esta permutación. El problema radica en determinar si existe algún ciclo de largo mayor que 50. En este caso, algún prisionero fallará en encontrar su propio número. Si todos los ciclos tienen largo menor que 50, entonces, todos los prisioneros serán exitosos. La probabilidad de que todos los ciclos tengan largo menor a 50 es un problema de combinatoria complicado, pero se puede demostrar que esta es la mejor estrategia posible y que brinda una probabilidad razonable de sobrevivencia, cercana al 30%, como hemos señalado.

Se invita al lector a trabajar este problema con una cantidad menor de prisioneros, alguna cantidad entre cinco y ocho, a identificar las permutaciones favorables y desfavorables y, de ser posible, a calcular la probabilidad de sobrevivencia.

El problema de la secretaria

El *problema de la secretaria* fue formulado en 1949 por Merrill M. Flood (1908-1991), un matemático norteamericano. Flood fue también, junto a Melvin Dresher, el creador del dilema del prisionero que analizamos anteriormente. La formulación del problema es la siguiente:

> *Una persona debe entrevistar a 100 candidatos para un único puesto de trabajo. Al realizar la entrevista, el entrevistador le asigna un puntaje al candidato e, inmediatamente, tiene que decidir si contratarlo o no. Esta es la única instancia en que puede contratar a esta persona. Una vez que pasa al candidato siguiente, no puede volver atrás.*
>
> *¿Cuál es la estrategia que maximiza las chances de seleccionar el mejor candidato?*

Recalcamos de partida que no hay un tope en el puntaje asignado al candidato. Si se le asigna una nota del 1 al 20, sería por supuesto razonable seleccionar el primer candidato con nota 20 o 19.

El problema de la secretaria cae en el ámbito de los problemas de **parada óptima**. Tenemos que elegir el instante en el cual detener cierto proceso para maximizar cierta "ganancia". Para decidir este instante, solo disponemos de la historia del proceso hasta el momento dado y no sabemos nada de los resultados futuros del mismo. Este enunciado es vago, pero una variedad de situaciones concretas cae dentro de esta categoría. Por ejemplo, supongamos que queremos comprar una propiedad. Las tasas de los créditos hipotecarios pueden variar bastante en el tiempo. Tenemos que decidir el instante en el cual contratar el crédito sin saber el valor de las tasas de interés en el futuro. Si supiéramos que están en alza, lo más sensato es comprar de inmediato. Por supuesto, nada excluye que las tasas bajen en un futuro cercano. Solo disponemos de los registros pasados para decidir.

La estrategia para seguir consiste en elegir un número **N** entre 1 y 100, por ejemplo, **N** = 50, y entrevistar a los 49 primeros candidatos. Ninguno de ellos se selecciona. Luego, se contrata al primer candidato a partir del número 50 en adelante cuya nota es mayor a la nota de todos los 49 primeros candidatos. Se puede demostrar que, con este método, se selecciona al mejor candidato con una probabilidad cercana a 35%.

En el párrafo anterior, 50 fue un número arbitrario. Podríamos haber decidido entrevistar un número diferente de candidatos. Si **N** = 60, entrevistamos a 59 candidatos, y la mejor nota de este grupo será posiblemente mayor que la mejor nota de los 49 primeros candidatos. Luego, el candidato seleccionado con **N** = 60 será mejor que con **N** = 50. Sin embargo, se puede demostrar que con **N** = 60 la probabilidad de seleccionar el mejor candidato cae a aproximadamente 31%. ¿A qué se debe esta aparente contradicción? El mejor candidato se distribuye uniformemente entre los 100 candidatos. Tiene una probabilidad positiva de estar entre los candidatos 50 y 59. Luego, al aumentar **N** de 50 a 60, también aumentamos las chances de incluir al mejor candidato en el grupo de aquellos que se desechan.

Por lo tanto, disminuyamos el valor de **N**. Al disminuir **N**, bajamos el "puntaje de corte", pero también bajamos la probabilidad de desechar automáticamente al mejor candidato. Con **N** = 40, la probabilidad de seleccionar el mejor candidato sube a 36,72%. Tomemos entonces un valor de **N** aun menor. Para **N** = 30, la probabilidad baja levemente a 35,89%.

Todo esto sugiere que debe existir un **N** óptimo, es decir, un **N** que maximiza la probabilidad de seleccionar el mejor candidato. Las probabilidades anteriores se obtuvieron mediante un cálculo explícito cuyos detalles no se escribieron; es una fórmula complicada que involucra herramientas básicas de cálculo universitario. Podemos probar con todos los números **N** entre 31 y 39 y quedarnos con el mejor valor. Resulta que este valor es 37, el cual otorga una probabilidad de 36,79% de seleccionar al mejor candidato.

Pero ¿por qué este misterioso valor **N** = 37? La respuesta radica en una constante omnipresente en las ciencias que se denota por la letra **e**. Su valor es aproximadamente igual a 2,7 y 100/2,7 ≈ 37.

Como ya lo indicamos, este problema puede resolverse con un cálculo explícito. Esta solución exacta puede a su vez aproximarse por una función cuyo valor máximo se puede encontrar usando herramientas del cálculo diferencial y es igual a **N/e**.

El número de candidatos se puede modificar. Si tenemos que entrevistar 200 candidatos, el mejor valor de **N** es 200/**e** ≈ 74. Si entrevistamos a 10.000 candidatos, entonces tenemos que tomar **N** = 10.000/**e** ≈ 3679.

El número *e*: las matemáticas abundan en números misteriosos que aparecen en multitudes de contextos dispares. Es el caso del famoso número π, cuya secuencia infinita de decimales no se repite jamás. Es también el caso de ***e***, conocido como la constante de Euler o de Néper, cuyo valor es aproximadamente igual a 2,72. Fue descubierta en el siglo XVII y se relaciona especialmente con los logaritmos. En 1737, Euler demostró que, al igual que π, esta constante es un número irracional, es decir, no se puede escribir como una fracción.

El juego de Penney

El juego de Penney es un juego de cara o sello muy sencillo que funciona de la manera siguiente:

El jugador A elige una secuencia de cara o sello de largo 3, por ejemplo, CSS o SCS, y se la muestra al jugador B. Acto seguido, el jugador B elige su propia secuencia.

Luego, se lanza una moneda hasta que aparece la secuencia elegida por alguno de los jugadores. El jugador cuya secuencia aparece primero gana.

¿Cuál es la mejor estrategia que seguir para el jugador B?

Lo curioso de este juego es que siempre existe una estrategia que le permite a B ganar con probabilidad mayor que ½. La estrategia es, por lo demás, muy sencilla y daremos algunos ejemplos de su funcionamiento.

Si la secuencia elegida por A es CSS, entonces B debe elegir CCS. Si A elige SCS, entonces B debe elegir SSC. En general, la elección del primer elemento de la secuencia de B debe ser el opuesto del segundo elemento de la secuencia de A. El segundo y tercer elemento elegidos por B deben coincidir con la primera y segunda elección de A.

El lector podrá verificar, estudiando los distintos casos posibles, que esta estrategia siempre es beneficiosa para B.

Hay algo profundamente contraintuitivo en este juego. No existe una estrategia mejor que todas las demás. Para cada elección de A, B puede elegir una secuencia que le prometa ganar con probabilidad mayor a ½. Por ejemplo, si A elige CSC, B tiene mayores chances de ganar eligiendo CCS. Pero si, a su vez, A elige CCS, B tendrá alta probabilidad de éxito escogiendo SCC.

El problema del asiento de avión

Nuestro problema final es muy fácil de formular, pero su solución, que queda a cargo del lector, requiere de mucho ingenio. Esta es su formulación clásica:

> *Los 300 pasajeros de un vuelo están a punto de abordar. El avión tiene exactamente 300 asientos. Cada pasajero tiene su ticket de embarque que indica el asiento donde deberá sentarse. Sin embargo, el primer pasajero que ingresa al avión pierde su ticket y, por lo tanto, no sabe qué asiento le corresponde. Algo abrumado, se sienta al azar en uno de los 300 asientos del avión.*
>
> *El siguiente pasajero que ingresa al avión se dirige a su asiento. Hay dos posibilidades: o bien su asiento está desocupado como debiera o está ocupado por el primer pasajero. En el primer caso, simplemente se sienta donde corresponde. En el segundo caso, se sienta al azar en alguno de los demás 299 asientos.*
>
> *Y así sucesivamente: cada pasajero que ingresa al avión se dirige a su asiento correspondiente. Si está desocupado, se sienta. Si, por el contrario, está ocupado por uno de los pasajeros anteriores, se sienta al azar en uno de los asientos restantes.*
>
> *¿Cuál es la probabilidad de que el último pasajero en ingresar al avión se siente en el puesto que le había sido asignado originalmente?*

Existe una solución sencilla y algo rutinaria a este problema que apela a una técnica conocida como inducción matemática. El lector que conoce este método puede intentarlo por esta vía.

Hay, sin embargo, una solución elemental e ingeniosa. Como única pista, dejamos la observación siguiente que deberá ser demostrada por el lector: al ingresar al avión, el último pasajero se encuentra con un único asiento desocupado. Hay solo dos posibilidades: o bien este asiento es el suyo o corresponde al asiento del primer pasajero.

BIBLIOGRAFÍA SELECCIONADA

Althoen, S.C., King, L., Schilling, K., "How long is a game of snakes and ladders?", *The Mathematical Gazette*, Vol. 77, No. 478, marzo 1993.

Arratia, R., Gordon, L., Waterman, M.S., "The Erdos-Renyi law in distribution, for coin tossing and sequence matching", *The Annals of Statistics*, Vol. 18, No. 2, junio 1990.

Barbeau, E., "Fallacies, flows and flimflaw", *The College Mathematics Journal*, Vol. 24, No. 2, marzo 1993.

Bennett, D., *Randomness*, Harvard University Press, 1998.

Berg, H., *Random walks in biology*, Princeton University Press, 1983.

Berg, H., *E. Coli in motion*, Springer, 2004.

Brin, S., Page, L., "The anatomy of a large-scale hypertextual web search engine", *Computer Networks and ISDN Systems*, Vol. 30, 1998.

Brown, R., "A brief account of microscopical observations made on the particles contained in the pollen of plants", *Philosophical Magazine*, Vol. 4, 1828.

Biswanger, K., Embrecht, P., "Longest runs in coin tossing", *Insurance: Mathematics and Economics*, Vol. 15, 1994.

David, F. N., *Games, gods and gambling*, Hafner Publishing Company, 1962.

Diaconis, P., Graham, R., *Magical mathematics*, Princeton University Press, 2012.

Duplantier, B., "Le mouvement brownien, divers et ondoyant", *Séminaire Poincaré*, Vol. 1, 2005.

Flajolet, P., Sedgewick, R., *Analytic combinatorics*, Cambridge University Press, 2009.

Gardner, M., "Problems involving questions of probability and ambiguity", columna Mathematical Games, *Scientific American*, Vol. 201, No. 4, 1959.

Gardner, M., "A fifth collection of brain teasers", columna Mathematical Games, *Scientific American*, Vol. 202, No. 2, 1960.

Gardner, M., *Hexaflexagons, probability paradoxes and the tower of Hanoi*, Cambridge University Press, 2008.

Gelman, A., Nolan, D., "You can load a die but you can't bias a coin", *The American Statistician*, Vol. 56, No. 4, noviembre 2002.

Gupta, P., Goel A., Lin, J., Sharma, A., Wang, D., Zadeh, R., "WTF: the who to follow service at Twitter", Proceedings of the 22nd international conference on World Wide Web, 2013.

Hill, T., "Knowing when to stop", *American Scientist*, Vol. 97, No. 2, 2009.

Kahnman, D., Tversky, A., *Evidential impact of base rates*, reporte técnico, 1981.

Keller, J., "The probability of heads", *The American Mathematical Monthly*, Vol. 93, No. 3, marzo 1986.

Kleinberg, J., "Authoritative sources in hyperlink environment", *Journal of the ACM*, Vol. 46, No. 5, 1999.

Lempel, R., Moran, S., "The stochastic approach for link-structure analysis (SALSA) and the TKC effect", *Computer Networks*, Vol. 33, 2000.

Oksendal, B., *Stochastic differential equations*, 5ª edición, Springer, 2000.

Ross, S., *A First course in probability*, 5ª edición, Prentice Hall, 1997.

Vos Savant, M., en: Ask Marilyn, *Parade Magazine*, 9 septiembre 1990.

Schilling, M., "The longest run of heads", *The College Mathematics Journal*, Vol. 21, No. 3, mayo 1990.

Székely, G., *Paradoxes in probability theory and mathematical statistics*, D. Reidel Publishing Company, 1986.

Winkler, P., *Mathematical mind-benders*, A. K. Peters Ltd., 2007.

Catalonia

Otras publicaciones

CINV
DeMente. El cerebro, un hueso duro de roer

MARCIAL ARREDONDO
Lógica. Elementos básicos

MARCIAL ARREDONDO
Aprendiendo a pensar. Manual de razonamiento para niños y jóvenes

MARCIAL ARREDONDO
Método de comprensión lectora. Entender lo que leemos. Comunicar lo que entendemos

PABLO ASTUDILLO
Manifiesto por la ciencia. Un nuevo relato para la ciencia en Chile

MARÍA TERESA RUIZ (COMP.)
Desde Chile: un cielo estrellado. Lecturas para fascinarse con la astronomía

BÁRBARA SILVA
Estrellas desde el San Cristóbal. La singular historia de un observatorio pionero en Chile

PERE CASTELLS
La cocina del futuro

www.catalonia.cl

El azar es el motor de una gran cantidad de fenómenos, desde la lotería hasta el movimiento de las bacterias. A veces, parecemos estar completamente a su merced y, otras veces, nos puede traer gratas sorpresas.

La teoría de probabilidades es la matemática del azar. Si bien no nos permite eliminar la incertidumbre del mundo que nos rodea, nos brinda una herramienta poderosa para entender sus leyes, tomar decisiones en situaciones inciertas e incluso sacar provecho de ellas. Hoy en día, la probabilidad está presente en los modelos matemáticos de la física, la biología, las finanzas e, incluso, en los algoritmos de las redes sociales.

Nuestro *Paseo por el azar* es un primer acercamiento a esta teoría. Recorreremos su historia, conoceremos a algunos de sus héroes y vislumbraremos algunas de sus aplicaciones actuales. Haciendo uso de matemáticas elementales, estudiaremos sus fundamentos y, en ocasiones, nos deleitaremos con sus paradojas inverosímiles.

ISBN 978-956-324-757-2

Gregorio R. Moreno Flores

UN PASEO POR EL AZAR

Catalonia