

4. DISTRIBUCIONES MUESTRALES Y PROBABILIDAD

DAVID BLACKWELL

La estadística aplicada se apoya en parte en la estadística teórica. La estadística ha avanzado no sólo gracias a personas enfrentadas a problemas prácticos, desde Florence Nightingale hasta R. A. Fisher y John Tukey, sino también gracias a personas cuyo principal interés han sido las matemáticas. David Blackwell (1919-) es el autor de algunas de las principales aportaciones al estudio matemático de la estadística.

Blackwell creció en Illinois, EE UU, donde se doctoró en Matemáticas a los 22 años de edad. En 1944 se incorporó al cuerpo docente de la Howard University en Washington, D.C. “En aquellos días, la ambición de cualquier universitario negro era poder llegar a ser profesor de la Howard University; era el mejor trabajo al que podías aspirar”, dice Blackwell. La sociedad cambió y, en 1954, Blackwell se convirtió en profesor de estadística de la University of California en Berkeley.

En Washington, D.C., donde había un activo grupo de estadísticos, el joven matemático Blackwell empezó a trabajar pronto en los aspectos matemáticos de la estadística. Blackwell exploró el comportamiento de los procedimientos estadísticos que, en vez de trabajar con una muestra fija, siguen obteniendo observaciones hasta que la información es suficiente para alcanzar una conclusión sólida. También descubrió nuevos aspectos de la inferencia considerándola como un juego en el cual la naturaleza compite en contra del estadístico. Los trabajos de Blackwell utilizan la teoría de la probabilidad, las matemáticas que describen el comportamiento del azar. Tenemos que seguir el mismo camino, aunque sólo durante una distancia corta. Este capítulo presenta, de manera más bien informal, las ideas de probabilidad necesarias para comprender los razonamientos de la inferencia.

4.1 Introducción

Los razonamientos de la inferencia estadística se basan en preguntar: “¿con qué frecuencia este método daría una respuesta correcta si lo utilizara muchas veces?”. La inferencia es más segura cuando obtenemos los datos a partir de muestras aleatorias o a partir de experimentos comparativos aleatorizados. La razón es que cuando utilizamos el azar para escoger a los individuos de una muestra o para asignar los sujetos de un experimento a los distintos tratamientos, las leyes de la probabilidad permiten responder a la pregunta: “¿qué ocurriría si lo repitiéramos muchas veces?”. El objetivo de este capítulo es entender lo que nos dicen las leyes de la probabilidad, pero sin entrar en las matemáticas de la teoría probabilística.

4.2 Aleatoriedad

¿Cuál es la media de los ingresos de los hogares estadounidenses? En EE UU, la Encuesta de Población Activa (EPA) constaba en 1997 de 50.000 hogares. La media de sus ingresos era $\bar{x} = 49.692$ dólares.¹ El valor 49.692 dólares describe la muestra, pero nosotros lo utilizamos para estimar la media de los ingresos de todos los hogares. Tenemos que tener bien claro si un valor describe a una muestra o a una población. He aquí la terminología que utilizaremos.

PARÁMETRO, ESTADÍSTICO

Un **parámetro** es un número que describe la población. En la práctica estadística el valor del parámetro no es conocido ya que no podemos examinar toda la población.

Un **estadístico** es un número que se puede calcular a partir de los datos de la muestra sin utilizar ningún parámetro desconocido. En la práctica, solemos utilizar un estadístico para estimar el parámetro desconocido.

¹U.S. Bureau of the Census, Current Population Report, P60-200, *Money Income in the United States*, 1997. Government Printing Office, Washington, D.C., 1998.

EJEMPLO 4.1. Ingresos de los hogares

Los ingresos medios de la muestra de la Encuesta de Población Activa (EPA) de EE UU era $\bar{x} = 49.692$ dólares. El número 49.692 es un *estadístico*, ya que describe una muestra en concreto. La población sobre la que la muestra trata de obtener conclusiones son los 103 millones de hogares de EE UU. El *parámetro* de interés, cuyo valor desconocemos, es la media de los ingresos de todos estos hogares norteamericanos. ■

Recuerda: los estadísticos proceden de muestras y los parámetros de poblaciones. Cuando sólo analizábamos datos, la distinción entre población y muestra no era importante. Sin embargo, ahora esta distinción es esencial. La notación que utilizemos debe reflejar esta diferencia.

Escribimos μ (la letra griega my) para indicar la **media poblacional**. Es un valor fijo cuyo valor es desconocido cuando utilizamos una muestra para hacer inferencias. La **media muestral** es la conocida $\bar{x}$, la media de observaciones de la muestra. $\bar{x}$ es un estadístico que casi seguro que hubiera tomado otro valor si hubiéramos escogido otra muestra de la misma población. La media $\bar{x}$ de una muestra o de un experimento es una estimación de la población de la media μ de la población muestreada.

*Media
poblacional
 μ
Media
muestral $\bar{x}$*

APLICA TUS CONOCIMIENTOS

A continuación, di si los números en negrita de los ejercicios del 4.1 al 4.3 son *parámetros* o *estadísticos*.

4.1. Los cojinetes de un lote de fabricación tienen **2,5003** centímetros (cm) de diámetro medio, que cumple con las condiciones fijadas por el comprador para la aceptación del lote. Un inspector escoge al azar 100 cojinetes del envío, que resultan tener un diámetro medio de **2,5009** cm. Como este valor excede del diámetro acordado por comprador y vendedor, el envío es rechazado erróneamente.

4.2. Una empresa de Los Ángeles que realiza estudios de mercado utiliza un aparato que marca al azar números de teléfono de esa ciudad. De los 100 primeros números marcados el **48%** no aparece en la guía telefónica. No es sorprendente, ya que el **52%** de los teléfonos de los Ángeles no están en la guía.

4.3. Una investigadora lleva a cabo un experimento comparativo aleatorizado con ratas jóvenes, con el fin de investigar los efectos de un compuesto tóxico en la comida. La investigadora alimenta al grupo de control con una dieta normal. El grupo experimental recibe una dieta con 2.500 partes por millón de una sustancia tóxica. Después de 8 semanas, el aumento de peso medio de las ratas es de 335 gramos en el grupo de control y de 289 gramos en el grupo experimental.

4.2.1 Concepto de probabilidad

¿Cómo es posible que la media $\bar{x}$ obtenida a partir de una muestra de unos pocos hogares de todos los del país, pueda ser una estimación precisa de μ ? Después de todo, una segunda muestra aleatoria obtenida en el mismo momento estaría formada por hogares distintos y, sin duda, daría un valor distinto de $\bar{x}$. A este hecho básico se le llama **variabilidad muestral**: el valor de un estadístico varía en un muestreo aleatorio repetido.

Para comprender por qué la variabilidad muestral no es fatal, debemos fijarnos en el comportamiento del azar. Un hecho importante es que **el comportamiento del azar es impredecible con pocas repeticiones pero presenta un comportamiento regular y predecible con muchas repeticiones**.

Lanza una moneda al aire o escoge una muestra aleatoria simple. A priori no se puede predecir el resultado, ya que variará cuando repitas el lanzamiento de la moneda o cuando obtengas la muestra. De todas formas, existe un comportamiento regular de los resultados, una regularidad que aparece de forma clara sólo después de muchas repeticiones. Este hecho remarcable es la base de la idea de probabilidad.

EJEMPLO 4.2. Lanzamiento de una moneda

Cuando lanzas una moneda al aire sólo hay dos resultados posibles, cara o cruz. La figura 4.1 muestra los resultados de lanzar una moneda 1.000 veces. Para cada lanzamiento, desde el primero hasta el último, hemos representado la proporción de lanzamientos que han dado cara hasta ese momento. El primer lanzamiento fue cara, por tanto, la proporción de caras empieza siendo 1. El segundo lanzamiento fue cruz. Después de dos lanzamientos, la proporción de caras se ha reducido a 0,5. Los siguientes tres lanzamientos dieron una cruz seguida de dos caras, por consiguiente, la proporción de caras después de cinco lanzamientos es $\frac{3}{5}$ o 0,6.

Variabilidad
muestral

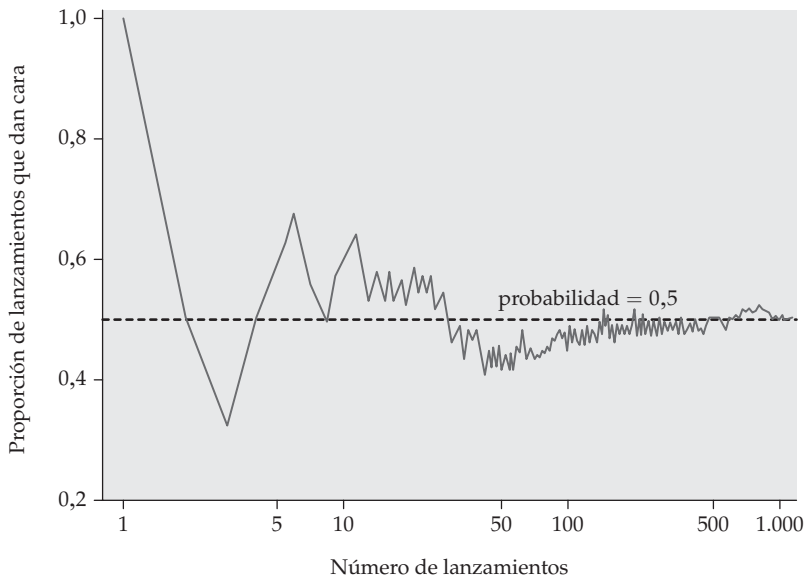


Figura 4.1. La proporción de lanzamientos de una moneda que dan cara cambia a medida que hacemos más lanzamientos. De todas formas, el valor de esta proporción se aproxima a 0,5, la probabilidad de cara.

La proporción de lanzamientos que dan cara es bastante variable al principio, pero posteriormente se estabiliza a medida que hacemos más y más lanzamientos. Llega un momento en que esta proporción se acerca a 0,5 y se mantiene en ese valor. Decimos que 0,5 es la *probabilidad* de que salga cara. La probabilidad 0,5 se ha dibujado como una línea horizontal discontinua en el gráfico. ■

“Aleatorio” en estadística no significa de “cualquier manera”, sino que se refiere a una clase de orden que únicamente aparece después de muchas repeticiones. La cara más impredecible de la aleatoriedad es nuestra experiencia del día a día: es difícil que veamos suficientes repeticiones de un mismo fenómeno aleatorio como para que observemos la regularidad que aparece después de muchas reiteraciones. Puedes ver la aparición de la regularidad en la figura 4.1. Después de muchas repeticiones, la proporción de lanzamientos que dan cara es 0,5. Esta es la idea intuitiva de probabilidad. Una probabilidad de 0,5 significa “ocurre la mitad de las veces después de un gran número de ensayos”.

Podíamos sospechar que una moneda tiene una probabilidad de 0,5 de que salga cara, solamente por el hecho de que las monedas tienen dos caras. Tal como ilustran los ejercicios 4.4 y 4.5, estas sospechas no son siempre correctas. La idea de probabilidad es empírica, es decir, se basa más en la experiencia que en la teoría. La probabilidad describe lo que ocurre después de muchísimos ensayos. Tenemos que observar muchas repeticiones para conocer una probabilidad. En el caso de lanzar una moneda, algunas personas tenaces han llevado a cabo millares de lanzamientos.

EJEMPLO 4.3. Algunos lanzamientos de monedas

El naturalista francés Count Buffon (1707-1788) lanzó al aire una moneda 4.040 veces. El resultado: 2.048 caras, o una proporción de caras de $\frac{2.048}{4.040} = 0,5069$.

Cerca del año 1900, el estadístico inglés Karl Pearson lanzó al aire una moneda 24.000 veces. El resultado: 12.012 caras, una proporción de 0,5005.

Mientras estuvo preso por los alemanes durante la Segunda Guerra Mundial, el matemático sudafricano John Kerrich lanzó 10.000 veces una moneda al aire. El resultado: 5.067 caras, una proporción de 0,5067. ■

ALEATORIEDAD Y PROBABILIDAD

Llamamos a un fenómeno **aleatorio** si los resultados individuales son inciertos y, sin embargo, existe una distribución regular de los resultados después de un gran número de repeticiones.

La **probabilidad** de cualquier resultado de un fenómeno aleatorio es la proporción de veces que el resultado se da después de una larga serie de repeticiones.

APLICA TUS CONOCIMIENTOS

4.4. Probabilidad de obtener cara. Sostén una moneda por el borde sobre una superficie plana con el dedo índice y dale un golpe con el mismo dedo de la otra mano, de tal manera que gire rápidamente hasta que finalmente caiga con la cara o la

cruz hacia arriba. Repite esto 50 veces y toma nota del número de caras. Estima la probabilidad de cara.

4.5. Más sobre la probabilidad de obtener cara. Es posible que supongas que es evidente que la probabilidad de obtener cara cuando se lanza una moneda es aproximadamente $\frac{1}{2}$ debido a que la moneda tiene dos caras. Esta suposición no siempre es cierta. En el ejercicio anterior, en vez de lanzar una moneda al aire, la hicimos girar sobre su borde —ello cambió la probabilidad de obtener cara—. Ahora vamos a ensayar otra variación. Sostén una moneda por su borde sobre una superficie plana y golpéala lateralmente con un dedo de la otra mano de manera que caiga la moneda. ¿Cuál es la probabilidad de que la moneda caiga con la cara hacia arriba? Repítelo, al menos 50 veces, para estimar la probabilidad de obtener cara.

4.2.2 Pensando en la aleatoriedad

Se puede observar que en el mundo real algunas cosas ocurren de forma aleatoria. El resultado de lanzar una moneda, el lapso de tiempo transcurrido entre emisiones de partículas de una fuente radioactiva o los sexos de cada uno de los componentes de una camada de ratas de laboratorio son ejemplos de fenómenos aleatorios. También lo son el resultado de una muestra aleatoria o de un experimento aleatorizado. La teoría de la probabilidad es la rama de las matemáticas que describe el comportamiento aleatorio. Por supuesto que nunca podemos observar la probabilidad de forma exacta. Siempre podríamos, por ejemplo, seguir lanzando una moneda al aire. La probabilidad matemática es una idealización basada en imaginar lo que ocurriría después de una serie infinita de repeticiones.

La mejor manera de comprender la aleatoriedad es observar un comportamiento aleatorio —no solamente la regularidad que aparece después de muchas repeticiones, sino también los resultados impredecibles obtenidos después de pocas repeticiones—. Lo puedes hacer no sólo con la ayuda de dispositivos como los de los ejercicios 4.4 a 4.8, sino también simulando (imitando) un determinado fenómeno aleatorio con un ordenador lo que te permite una exploración más rápida. Los ejercicios 4.11 a 4.13 sugieren algunas simulaciones de un comportamiento aleatorio. Cuando explores la aleatoriedad recuerda que:

- Tienes que tener una larga serie de ensayos **independientes**. Es decir, el resultado de un ensayo no debe influir sobre el resultado de otro. Imagina que en un casino el crupier hace trampas de manera que pueda parar la ruleta cuando desee y de esta manera conseguir una determinada

Independencia

proporción de “rojos”. Los resultados de estos ensayos (hacer girar la ruleta) no son independientes.

- La idea de probabilidad es empírica. Las simulaciones con ordenador parten de unas probabilidades predeterminadas e imitan un comportamiento aleatorio. Sin embargo, en el mundo real solamente podemos estimar las probabilidades observando muchos resultados de un determinado fenómeno.
- De todas formas, las simulaciones con ordenador son muy útiles, ya que necesitamos los resultados de muchos ensayos. En situaciones como el lanzamiento de una moneda, la proporción de un determinado resultado a menudo exige centenares de repeticiones para poder estimar su probabilidad de aparición. Los dispositivos propuestos en los ejercicios no permiten operar muy deprisa. Pocas repeticiones sólo permiten toscas estimaciones de la probabilidad.

RESUMEN DE LA SECCIÓN 4.2

Los resultados de un fenómeno aleatorio no se pueden predecir, de todas formas después de muchas repeticiones la distribución de resultados es regular.

La **probabilidad** de un suceso es la proporción de veces que ocurre después de muchas repeticiones de un determinado fenómeno aleatorio.

EJERCICIOS DE LA SECCIÓN 4.2

4.6. Dígitos aleatorios. La tabla de dígitos aleatorios (tabla B) se elaboró con un programa de aleatorización que da a cada dígito una probabilidad igual a 0,1 de ser 0? ¿Qué proporción de los 200 primeros dígitos de la tabla son ceros? Esta proporción es una estimación de la verdadera probabilidad, basada en 200 repeticiones. En este caso se sabe que la verdadera probabilidad es 0,1.

4.7. ¿Cuántos lanzamientos hasta obtener cara? La experiencia demuestra que cuando lanzas una moneda, la probabilidad de obtener cara (la proporción de caras después de muchas repeticiones) es $\frac{1}{2}$. Supón que lanzas al aire una moneda muchas veces, hasta que obtienes cara. ¿Cuál es la probabilidad de que la primera cara aparezca en un lanzamiento impar (1, 3, 5, etc.)? Para saberlo, repite este experimento 50 veces y apunta el número de lanzamientos que has llevado a cabo en cada uno de ellos.

(a) A partir de tus experimentos, estima la probabilidad de obtener cara en el primer lanzamiento. ¿Cuál debería ser el valor de esta probabilidad?

(b) A partir de tus resultados, estima la probabilidad de que la primera cara aparezca en un lanzamiento impar.

4.8. Lanzamiento de una chincheta. Lanza una chincheta al aire 100 veces sobre una superficie plana. ¿Cuántas veces cayó con la punta hacia arriba? ¿Cuál es la probabilidad aproximada de que caiga con la punta hacia arriba?

4.9. Trío. Supón que lees en un libro de póquer que la probabilidad de obtener un trío cuando se reparten 5 cartas a cada jugador es $\frac{1}{50}$. Explica en un lenguaje sencillo lo que esto significa.

4.10. La probabilidad es una medida de la posibilidad que tiene un suceso de ocurrir. Asocia cada una de las siguientes probabilidades con cada una de las afirmaciones presentadas. (En general, la probabilidad es una medida más exacta de las posibilidades de un fenómeno que una afirmación verbal.)

0 0,01 0,3 0,6 0,99 1

(a) Este suceso es imposible. No puede ocurrir nunca.

(b) Este suceso es seguro. Ocurrirá en todas las repeticiones de este fenómeno aleatorio.

(c) Este suceso es muy difícil de que ocurra, pero tendrá lugar de vez en cuando en una larga secuencia de repeticiones.

(d) Este suceso ocurrirá la mayoría de las veces.

4.11. Tiros libres encestados. Una jugadora de baloncesto, después de toda la temporada, encesta como media aproximadamente la mitad de los tiros libres. Utiliza un programa informático para simular el lanzamiento de 100 tiros libres independientes de un jugador que tiene una probabilidad de encestar de 0,5. (En muchos programas informáticos el procedimiento clave para esta simulación son las “pruebas de Bernoulli”. Equivalen a pruebas independientes con dos resultados posibles. Nuestros resultados son “Encestar” y “Fallar”.)

(a) De 100 tiros libres, ¿qué porcentaje encesta?

(b) Examina la secuencia de encestes y fallos. ¿De cuántos resultados se compone la serie más larga de encestes? ¿Y la de fallos? (A menudo, las secuencias de resultados aleatorios presentan series de resultados iguales más largas de lo que imaginaríamos.)

4.12. Jugar en casa. Un estudio sobre las ventajas de jugar a baloncesto en casa, halló que entre los años 1969 y 1989 en la liga de baloncesto se ganaron el 63% de los partidos jugados en casa.² ¿Utilizarías los resultados de este estudio para asignar un valor igual a 0,63 a la probabilidad de ganar en casa? Justifica tu respuesta.

4.13. Simulación de una encuesta de opinión. Una reciente encuesta de opinión mostró que el 73% de las mujeres casadas está de acuerdo con que sus esposos contribuyen el mínimo imprescindible a las tareas del hogar. Supón que esto sea cierto. Si se escoge al azar una mujer casada, la probabilidad de que esté de acuerdo con que su esposo contribuye el mínimo imprescindible a las tareas domésticas es 0,73. Utiliza un programa informático para simular la elección de muchas mujeres de forma independiente. (En muchos programas el procedimiento clave para esta simulación son las “pruebas de Bernoulli”. Equivale a pruebas independientes con dos resultados posibles. Nuestros resultados son “Está de acuerdo” y “No está de acuerdo”).

(a) Simula la obtención al azar de 20 mujeres, luego de 80 y después de 320. ¿Qué proporción de mujeres está de acuerdo en cada caso? Creemos (pero debido a la variación del azar no estamos seguros) que después de muchas repeticiones, la proporción estará cerca de 0,73.

(b) Simula 10 veces la obtención al azar de 20 mujeres. Apunta el porcentaje de mujeres en cada experimento que “está de acuerdo”. Luego simula 10 veces la obtención de 320 mujeres al azar y apunta otra vez los porcentajes. ¿Qué conjunto de 10 resultados es menos variable? Creemos que los resultados de 320 ensayos serán más fáciles de predecir (menos variables) que los resultados de 20 repeticiones. Así se pone de manifiesto la regularidad después de muchas repeticiones.

4.3 Modelos de probabilidad

En capítulos anteriores vimos modelos matemáticos para relaciones lineales (en forma de ecuación para una recta) y para algunas distribuciones de datos (en forma de curvas de densidad normales). Ahora vamos a ver una descripción matemática, o modelo, para la aleatoriedad. Para ver cómo procedemos, piensa en

²W. Hurley, “What sort of tournament should the World Series be?”, *Chance*, 6, nº 2, 1993, págs. 31-33.

primer lugar, en un fenómeno aleatorio muy simple; el lanzamiento de una moneda al aire. Cuando la lanzamos, a priori no podemos prever lo que saldrá. ¿Qué es lo que conocemos? Lo único que sabemos es que puede salir cara o cruz. Creemos que cada uno de estos fenómenos tiene una probabilidad de $\frac{1}{2}$ de ocurrir. La descripción del lanzamiento de una moneda consta de dos partes:

- La lista de resultados posibles.
- La probabilidad de cada resultado.

Este tipo de descripción es la base de cualquier modelo de probabilidad. La terminología básica que utilizaremos es la siguiente:

MODELOS DE PROBABILIDAD

El **espacio muestral** S de un fenómeno aleatorio es el conjunto de todos sus resultados posibles.

Un **suceso** es cualquier resultado o conjunto de resultados de un fenómeno aleatorio. Es decir, un suceso es un subconjunto del espacio muestral.

Un **modelo de probabilidad** es una descripción matemática de un fenómeno aleatorio. Consta de dos partes: un espacio muestral S y un procedimiento de asignación de probabilidades a los sucesos.

El espacio muestral S puede ser muy simple o muy complejo. Cuando lanzamos una vez una moneda, sólo hay dos resultados posibles: cara y cruz. El espacio muestral es $S = \{H, T\}$. Si escogemos una muestra aleatoria de 50.000 hogares, como en la Encuesta de Población Activa (EPA) de EE UU, el espacio muestral contiene todas las posibilidades de escoger 50.000 de los 103 millones de hogares de este país. Esta S es extremadamente grande. Cada componente de S es una posible muestra, lo que explica el término *espacio muestral*.

EJEMPLO 4.4. Lanzamiento de dos dados

Lanzar dos dados es una forma habitual de perder dinero en los casinos. Cuando lanzamos y nos fijamos de forma ordenada en los resultados de las caras superiores (primer dado, segundo dado) podemos obtener 36 resultados posibles. La figura 4.2 muestra estos resultados, que forman un espacio muestral S . “Obtener un 5” es un suceso, llámalo A , que está formado por cuatro de los 36 resultados posibles:

$$A = \left\{ \begin{array}{|c|c|} \hline \bullet & \bullet\bullet \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \bullet & \bullet\bullet\bullet \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \bullet\bullet & \bullet \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \bullet\bullet & \bullet\bullet\bullet \\ \hline \end{array} \right\}$$

Los jugadores sólo se fijan en la suma de los resultados de las caras superiores. El espacio muestral resultante de lanzar dos dados y sumar los resultados de las caras superiores es

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

Comparando esta S con la figura 4.2 nos recuerda que podemos cambiar S cambiando la descripción detallada del fenómeno aleatorio que consideramos. ■

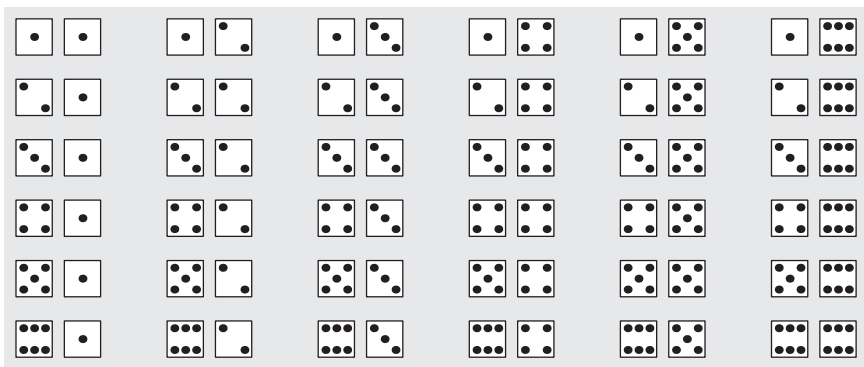


Figura 4.2. Los 36 resultados posibles del lanzamiento de dos dados.

APLICA TUS CONOCIMIENTOS

4.14. En cada una de las siguientes situaciones, describe el espacio muestral S del fenómeno aleatorio considerado. En algunos casos tienen varias posibilidades al definir S .

- (a) Se siembra una semilla. La semilla o bien germina o bien se muere.
- (b) Un paciente con un tipo de cáncer incurable se somete a un nuevo tratamiento. La variable respuesta es el tiempo que sobrevive el paciente después del tratamiento.
- (c) Un estudiante se matricula un curso de estadística. Al final del curso recibe la nota que ha obtenido.
- (d) Un jugador de baloncesto lanza cuatro tiros libres. Registras la secuencia de aciertos y de fallos.
- (e) Un jugador de baloncesto lanza cuatro tiros libres. Registras el número de canastas.

4.15. En cada una de las siguientes situaciones, describe el espacio muestral S del fenómeno aleatorio considerado. En algunos casos tienes varias posibilidades al definir S , especialmente al determinar los valores mayores y menores de S .

- (a) Escoge al azar un estudiante de tu clase. Pregúntale cuánto tiempo pasó estudiando en las últimas 24 horas.
- (b) El *Physicians' Health Study* pidió a 11.000 médicos que tomaran una aspirina cada dos días y observó cuántos de ellos sufrieron ataques al corazón durante un periodo de 5 años.
- (c) En una prueba sobre un nuevo embalaje para huevos, dejas caer una caja embalada desde una altura de medio metro y haces un recuento del número de huevos rotos.
- (d) Escoge al azar un estudiante de tu clase y pregúntale cuánto dinero lleva encima.
- (e) Un investigador sobre la nutrición alimenta a una rata joven con una nueva dieta. La variable respuesta es la ganancia de peso (en gramos) al cabo de 8 semanas.

4.3.1 Reglas de la probabilidad

La probabilidad de cualquier resultado —por ejemplo, obtener un cinco cuando se lanzan dos dados— sólo se puede hallar lanzando dos dados muchas veces, y

además sólo de forma aproximada. Entonces, ¿cómo podemos describir la probabilidad matemáticamente? Más que intentar dar probabilidades exactas, empezaremos por presentar algunas condiciones que se deben cumplir para poder asignar probabilidades. Estas condiciones surgen de la idea de probabilidad como “la proporción de veces que ocurre un determinado suceso después de muchas repeticiones”.

1. **Cualquier probabilidad es un número entre 0 y 1.** Cualquier proporción es un número entre 0 y 1, por tanto, cualquier probabilidad también es un número entre 0 y 1. Un suceso con una probabilidad 0 no ocurre nunca, mientras que un suceso con una probabilidad 1 ocurre en todas las repeticiones. Un suceso con una probabilidad de 0,5 ocurre, después de muchas repeticiones, la mitad de las veces.
2. **La probabilidad de todos los resultados posibles, considerados conjuntamente, tiene que ser 1.** Debido a que en cada repetición siempre obtenemos algún resultado, la suma de las probabilidades de todos los resultados posibles tiene que ser exactamente 1.
3. **La probabilidad de que un suceso no ocurra es 1 menos la probabilidad de que este suceso ocurra.** Si un suceso ocurre en, digamos, el 70% de las repeticiones, deja de ocurrir en el 30% de las mismas. La probabilidad de que un suceso ocurra y la probabilidad de que no ocurra siempre llega hasta el 100%, es decir, es 1.
4. **Si dos sucesos no tienen resultados en común, la probabilidad de que ocurra alguno de los dos es la suma de sus respectivas probabilidades.** Si un suceso ocurre el 40% de todas las repeticiones y otro suceso diferente ocurre el 25% de todas las repeticiones, y los dos sucesos no se pueden producir simultáneamente, entonces la probabilidad de que ocurra uno de ellos es del 65%, ya que $40\% + 25\% = 65\%$.

Podemos utilizar la notación matemática para describir de forma más precisa lo que hemos planteado en los puntos anteriores. Si A es un suceso, indicamos su probabilidad como $P(A)$. He aquí las condiciones que debe cumplir una probabilidad en un lenguaje formal. Cuando utilices estas condiciones, recuerda que no son más que otra forma de describir lo que acabamos de ver de forma intuitiva sobre lo que ocurre con las proporciones después de muchas repeticiones.

REGLAS DE LA PROBABILIDAD

Regla 1. La probabilidad $P(A)$ de cualquier suceso A cumple que

$$0 \leq P(A) \leq 1$$

Regla 2. Si S es el espacio muestral de un modelo de probabilidad, entonces $P(S) = 1$.

Regla 3. Para cualquier suceso A ,

$$P(\text{no ocurra } A) = 1 - P(A)$$

Regla 4. Dos sucesos A y B son **disjuntos** si no tienen resultados en común, es decir, no pueden ocurrir nunca de forma simultánea. Si A y B son disjuntos,

$$P(A \text{ o } B) = P(A) + P(B)$$

Esta es la **regla de la suma de sucesos disjuntos**.

EJEMPLO 4.5. Estado civil de mujeres jóvenes

Escoge al azar una mujer de entre 25 y 29 años, y pregúntale su estado civil. “Al azar” significa que damos a todas las mujeres jóvenes las mismas posibilidades de ser una de las escogidas. Es decir, escogemos una muestra aleatoria simple de tamaño 1. La probabilidad de cualquier estado civil no es más que la proporción de mujeres de entre 25 y 29 años que tienen este estado civil en la población —si escogiéramos muchas mujeres, esta sería la proporción que obtendríamos—. He aquí el modelo de probabilidad:

Estado civil	Soltera	Casada	Viuda	Divorciada
Probabilidad	0,353	0,574	0,002	0,071

Cada probabilidad toma un valor entre 0 y 1. La suma de las probabilidades es 1, ya que todos estos resultados constituyen el espacio muestral S .

La probabilidad de que la mujer que hemos escogido no esté casada es, según la Regla 3,

$$\begin{aligned} P(\text{no casada}) &= 1 - P(\text{casada}) \\ &= 1 - 0,574 = 0,426 \end{aligned}$$

Es decir, si el 57,4% están casadas, entonces las restantes 42,6% no lo están.

“Nunca casadas” y “divorciadas” son sucesos disjuntos, ya que ninguna mujer puede simultáneamente no haberse casado nunca y estar divorciada. En consecuencia, la regla de la adición establece que

$$\begin{aligned} P(\text{nunca casada o divorciada}) &= P(\text{nunca casada}) + P(\text{divorciada}) \\ &= 0,353 - 0,071 = 0,424 \end{aligned}$$

Es decir, el 42,4% de las mujeres de este grupo de edad o nunca se casaron o se divorciaron. ■

EJEMPLO 4.6. Probabilidades cuando se lanzan dos dados

La figura 4.2 muestra los 36 resultados posibles de lanzar dos dados. ¿Qué probabilidades debemos asignar a estos resultados?

Los dados de los casinos se fabrican con mucho cuidado. Las marcas de los puntos no se vacían, lo que daría diferente peso a las distintas caras, sino que se rellenan con un plástico transparente que tiene la misma densidad que el plástico utilizado para construir el dado. Para este tipo de dados es razonable asignar la misma probabilidad a los 36 resultados de la figura 4.2. Debido a que los 36 resultados considerados conjuntamente deben tener una probabilidad de 1 (Regla 2), cada resultado debe tener una probabilidad de $\frac{1}{36}$.

En general, los jugadores se interesan por la suma de las caras superiores. ¿Cuál es la probabilidad de obtener un 5? Debido a que el suceso “obtener un 5” está formado por los cuatro resultados mostrados en el ejemplo 4.4, la regla de la adición (Regla 4) establece que su probabilidad es

$$P = (\text{Obtener un 5}) = P\left(\begin{array}{|c|c|} \hline \bullet & \bullet \bullet \\ \hline \end{array}\right) + P\left(\begin{array}{|c|c|} \hline \bullet \bullet & \bullet \\ \hline \end{array}\right) + P\left(\begin{array}{|c|c|} \hline \bullet & \bullet \bullet \\ \hline \end{array}\right) + P\left(\begin{array}{|c|c|} \hline \bullet \bullet & \bullet \\ \hline \end{array}\right)$$

¿Cuál es la probabilidad de obtener un 7? En la figura 4.2 encontrarás los seis resultados para los cuales la suma de puntos es 7. La probabilidad es $\frac{6}{36}$, o aproximadamente 0,167. ■

APLICA TUS CONOCIMIENTOS

4.16. Ascenso social. Un sociólogo que estudia la movilidad social en Dinamarca halla que la probabilidad de que un hijo de padres de clase baja siga en dicha clase es de 0,46. ¿Cuál es la probabilidad de que el hijo ascienda a una clase más alta?

4.17. Causas de muerte. El Gobierno asigna a una sola causa las muertes que suceden en el país. Los datos muestran que la probabilidad de que una muerte escogida al azar se deba a una enfermedad cardiovascular es de 0,45 y de 0,22 que se deba al cáncer. ¿Cuál es la probabilidad de que una muerte se deba a una enfermedad cardiovascular o al cáncer? ¿Cuál es la probabilidad de que la muerte se deba a otra causa?

4.18. ¿Contribuyen los esposos a las tareas domésticas? El *New York Times* dio a conocer los resultados de una encuesta que realizó a una muestra aleatoria de 1.025 mujeres. A las mujeres casadas de la muestra se les preguntó si sus esposos contribuían a las tareas domésticas. He aquí los resultados:

Resultado	Probabilidad
Hace más de lo justo	0,12
Hace lo justo	0,61
Hace menos de lo justo	?

Estas proporciones son probabilidades de un fenómeno aleatorio que consiste en escoger una mujer casada al azar y preguntarle su opinión.

(a) ¿Cuál debería ser la probabilidad de que una mujer escogida al azar diga que su esposo hace menos de lo justo? ¿Por qué?

(b) El suceso “creo que mi esposo hace al menos lo justo” comprende dos resultados. ¿Cuál es su probabilidad?

4.3.2 Asignación de probabilidades: número finito de resultados

Los ejemplos 4.5 y 4.6 ilustran una manera de asignar probabilidades a los sucesos: asigna una probabilidad a cada uno de los resultados posibles, luego suma estas probabilidades para hallar la probabilidad de cualquier suceso. Para que este proceder cumpla las reglas de la probabilidad, la probabilidad de todos los resultados individuales tiene que sumar exactamente 1.

PROBABILIDADES EN UN ESPACIO MUESTRAL FINITO

Asigna una probabilidad a cada resultado individual. Estas probabilidades deben ser números entre 0 y 1. Su suma debe ser 1.

La probabilidad de cualquier suceso es la suma de las probabilidades de los resultados que lo constituyen.

EJEMPLO 4.7. Dígitos aleatorios

La tabla de dígitos aleatorios situada al final del libro se creó mediante un programa informático que generaba dígitos al azar entre 0 y 9. Si generamos un dígito, el espacio muestral es

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Una aleatorización correcta da a cada uno de los resultados las mismas posibilidades de ser uno de los elegidos. Debido a que la probabilidad global debe ser 1, la probabilidad de cada uno de los 10 resultados posibles debe ser $\frac{1}{10}$. Esta asignación de probabilidades a resultados individuales se puede resumir en la tabla siguiente:

Resultado	0	1	2	3	4	5	6	7	8	9
Probabilidad	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

La probabilidad del suceso de que ha salido un número impar es

$$P(\text{número impar}) = P(1) + P(3) + P(5) + P(7) + P(9) = 0,5$$

La asignación de probabilidades cumple todas nuestras reglas. Por ejemplo, podemos hallar la probabilidad de que salga un número par utilizando la Regla 3:

$$\begin{aligned}
 P(\text{número par}) &= P(\text{no sea un número impar}) \\
 &= 1 - P(\text{número impar}) \\
 &= 1 - 0,5 = 0,5
 \end{aligned}$$

Comprueba que obtienes el mismo resultado sumando las probabilidades de todos los resultados pares. ■

APLICA TUS CONOCIMIENTOS

4.19. Lanzamiento de un dado. La figura 4.3 muestra algunas asignaciones de probabilidad a las seis caras de un dado. Sólo podemos saber qué asignación de probabilidades es *correcta* lanzando el dado muchas veces. De todas formas, algunas de las asignaciones de probabilidad no son *admisibles*. Es decir, no cumplen las reglas. ¿Qué asignaciones de probabilidad son admisibles? En el caso de los modelos inadmisibles, explica la razón.

Resultado	Probabilidad			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	1/7	1/3	1/3	1
	1/7	1/6	1/6	1
	1/7	1/6	1/6	2
	1/7	0	1/6	1
	1/7	1/6	1/6	1
	1/7	1/6	1/6	2

Figura 4.3. Cuatro asignaciones de probabilidad a las seis caras de un dado, para el ejercicio 4.19.

4.20. Notas de alumnos de secundaria. Selecciona al azar a un estudiante universitario de primer curso y pregúntale qué posición ocupaba, según las notas obtenidas, en secundaria. He aquí las probabilidades de ocupar una determinada posición en secundaria basadas en una gran encuesta a estudiantes de secundaria:

Resultado	Primer 20%	Segundo 20%	Tercer 20%	Cuarto 20%	Quinto 20%
Probabilidad	0,41	0,23	0,29	0,06	0,01

(a) ¿Cuál es la suma de estas probabilidades? ¿Por qué crees tiene este valor?

(b) ¿Cuál es la probabilidad de que un estudiante universitario de primer curso escogido aleatoriamente no estuviera entre el 20% de los estudiantes con mejores notas de su clase en la escuela de secundaria?

(c) ¿Cuál es la probabilidad de que un estudiante de primer curso estuviera entre el 40% de los estudiantes con mejores notas en su clase de la escuela de secundaria?

4.21. Tipos de sangre. La sangre humana se puede clasificar en 4 grupos: O, A, B o AB. La distribución de los grupos varía un poco según la raza. He aquí las probabilidades de que una persona de raza negra escogida al azar en EE UU pertenezca a uno de los grupos posibles:

Grupo sanguíneo	O	A	B	AB
Probabilidad	0,49	0,27	0,20	?

(a) ¿Cuál es la probabilidad de que la persona escogida al azar pertenezca al grupo AB? ¿Por qué?

(b) María pertenece al grupo B y puede recibir sangre de los grupos O y B. ¿Cuál es la probabilidad de que un estadounidense de raza negra, escogido al azar, pueda donar sangre a María?

4.3.3 Asignación de probabilidades: intervalos de resultados

Supón que queremos escoger un número al azar entre 0 y 1, de manera que el resultado pueda ser cualquier número entre 0 y 1. Esto se puede hacer mediante un programa generador de números aleatorios. El espacio muestral ahora es todo el intervalo de números entre 0 y 1:

$$S = \{\text{todos los números entre 0 y 1}\}$$

Para abreviar llamamos al resultado del generador de números aleatorios Y . ¿Cómo podemos asignar probabilidades a sucesos como $\{0,3 \leq Y \leq 0,7\}$? De la misma manera que cuando seleccionamos un dígito aleatorio, queremos que todos los resultados posibles tengan las mismas posibilidades. Sin embargo, no podemos asignar probabilidades a cada valor individual de Y y luego sumarlos, ya que existe un número infinito de resultados posibles.

Utilizaremos un nuevo procedimiento para asignar probabilidades directamente a los sucesos —como *áreas por debajo una curva de densidad*—. El área por debajo de cualquier curva de densidad es 1 y corresponde a una probabilidad total igual a 1.