Probabilidad

1. Experimentos aleatorios. Probabilidad

Imaginemos que lanzamos un dado. El resultado del lanzamiento puede ser cualquiera de los números comprendidos entre 1 y 6, y, en principio, no se puede hacer ninguna predicción sobre cuál de esos resultados va a obtenerse. Si se realizan n lanzamientos, el número de veces que se obtienen los diversos resultados, es decir, sus **frecuencias absolutas** son n_1, n_2, n_3, n_4, n_5 y n_6. Se llama **frecuencia relativa** de cada resultado a la relación entre su frecuencia absoluta y n. Por ejemplo, la frecuencia relativa del 3 es $f_3 = n_3/n$. Además, se cumple que

$$n_1 + n_2 + n_3 + n_4 + n_5 + n_6 = n.$$

Dividiendo por n la igualdad anterior, se deduce que la suma de las frecuencias relativas de todos los resultados es igual a 1:

$$f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 1.$$

Cabe esperar que cuando el número de lanzamientos sea muy grande, las frecuencias relativas de todos los resultados sean aproximadamente iguales y, por consiguiente, tomen valores cercanos a $1/6$. Si no sucede esto, si por ejemplo, en la mitad de los lanzamientos se obtiene 5, habría que pensar que el dado es defectuoso o está trucado.

El lanzamiento de un dado, la extracción de una carta de una baraja, el lanzamiento de una moneda, etc., son ejemplos de lo que se conocen como **experimentos aleatorios** y se caracterizan por las siguientes propiedades:

1. Cuando se realiza el experimento una sola vez, no es posible hacer ninguna predicción sobre cuál de los posibles resultados se va a obtener.
2. Si el experimento se realiza un gran número de veces, es posible hacer predicciones sobre la proporción de veces que se va a obtener un determinado resultado, es decir, sobre su frecuencia relativa.

Se llama **probabilidad de un resultado** en un experimento aleatorio a la frecuencia relativa de ese resultado cuando el experimento se repite un gran número de veces (cuando dicho número tiende a infinito).

Puesto que las frecuencias relativas de los resultados de un experimento (sea aleatorio o no) suman 1, se deduce que la suma de las probabilidades de todos los resultados posibles debe ser igual a 1.

2. Cálculo de probabilidades

Dado que las probabilidades son las frecuencias relativas de los distintos resultados cuando el experimento se repite un gran número de veces, un modo de calcular la probabilidad de un resultado consiste en repetir el experimento muchas veces y medir la frecuencia relativa del resultado. A menudo se asignan probabilidades de esta manera. Así, cuando se dice que la probabilidad de supervivencia a determinada enfermedad es del 90 %, se está diciendo que la frecuencia relativa del resultado que consiste en superar la enfermedad en los casos observados, es del 90 %.

En otras muchas ocasiones se asignan las probabilidades atendiendo a la simetría del problema. Un dado presenta una simetría tal que no hay razón para asignar mayor probabilidad a una cara que a otra. Si los 6 resultados tienen la misma probabilidad y la suma de todas ellas debe ser igual 1, está claro que cada posible resultado debe tener una probabilidad igual a 1/6.

Se llama **espacio muestral** de un experimento al conjunto de todos los resultados que puede tener el experimento y se suele representar mediante la letra E. Por ejemplo, el espacio muestral del experimento que consiste en lanzar dos monedas y ver el resultado obtenido en cada una de ellas sería:

$$E = \{CC, CX, XC, XX\},$$

donde se ha representado por C el resultado "*cara*" y por X el resultado *"cruz"*. Si el experimento consiste en lanzar las dos monedas y contar el número de caras obtenido, el espacio muestral, dado que pueden obtenerse $0, 1$ ó 2 caras, es:

$$E = \{0, 1, 2\}.$$

Para calcular probabilidades hay que pensar en un espacio muestral que tenga la mayor simetría posible y, si cabe, que todos los resultados tengan la misma probabilidad. Una vez asignada la probabilidad a los resultados simples del espacio muestral, pueden calcularse probabilidades para resultados más complejos, como por ejemplo del tipo "*obtener un número primo al lanzar dos dados y sumar los puntos de cada uno*".

Cuando todos los resultados de un espacio muestral tienen la misma probabilidad, el espacio se llama **equiprobable**. La probabilidad de un resultado en un espacio equiprobable es igual a 1 dividido por el número de resultados posibles. Por ejemplo, la probabilidad de sacar el as de oros en una extracción aleatoria de una carta de una baraja española es 1/40 porque hay 40 resultados posibles y todos tienen la misma probabilidad. En el experimento de lanzar dos monedas, el espacio muestral $E = \{CC, CX, XC, XX\}$ es equiprobable y la probabilidad de cada resultado es 1/4. Sin embargo, el espacio muestral del número de caras obtenido $E = \{0, 1, 2\}$ no lo es. En general, siempre que sea posible, se tratará de diseñar el experimento de forma que el espacio muestral sea equiprobable.

Ejemplo 1 El espacio muestral al lanzar una moneda 3 veces y anotar los resultados de cada lanzamiento es

$$E = \{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}.$$

Si en cada lanzamiento la probabilidad de obtener "*cara*" es la misma que la de obtener "*cruz*", los 8 resultados de este espacio muestral tienen la misma probabilidad: 1/8.

Ejemplo 2 Si el experimento consistente en lanzar repetidamente una moneda y contar el número de lanzamientos efectuados hasta que sale "*cara*" por primera vez, el espacio muestral no es equiprobable y será:

$$E = \{1, 2, 3, \ldots\}.$$

---

3. Sucesos

Supóngase que se ha establecido el espacio muestral de forma que se conocen las probabilidades de todos los resultados que lo componen.

Por ejemplo, sea el experimento que consiste en lanzar dos dados. Se puede tomar como espacio muestral el conjunto:

$$\begin{aligned} E =\{&11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34, 35, 36,\\ &41, 42, 43, 44, 45, 46, 51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}, \end{aligned}$$

donde la primera cifra de cada número indica el resultado del primer dado y la segunda cifra el resultado del segundo.

El espacio así definido es equiprobable, dado que, por ejemplo, no debería ser diferente la probabilidad de los resultados 23 y 24 si en el segundo dado la probabilidad de sacar 3 es la misma que la de sacar 4. Razonando de esta manera, se llega a la conclusión de que los 36 resultados tienen la misma probabilidad y por tanto, la probabilidad de uno cualquiera de ellos es 1/36.

Los resultados que forman el espacio muestral se llaman **sucesos elementales**. El espacio muestral se ha construido de tal manera que la probabilidad de estos sucesos sea conocida. Sin embargo, a veces es preciso calcular probabilidades de resultados más complejos, por ejemplo, "*obtener suma* 10 *al lanzar dos dados*". Para resolver este problema, primero hay que definir el concepto de suceso.

Se llama **suceso** a un subconjunto del espacio muestral, es decir, a un conjunto formado por resultados del espacio muestral. Por ejemplo el suceso $A =$ "sacar suma 10 al lanzar dos dados" se corresponde con el subconjunto $A = \{46, 55, 64\}$ formado por 3 resultados elementales.

En un experimento aleatorio, un suceso se cumple cuando se obtiene como resultado alguno de los que forman parte del suceso. En el ejemplo anterior, se obtendrá suma 10 si al lanzar los dos dados resulta 6 en el primer dado y 4 en el segundo, 5 en los dos o 4 en el primer dado y 6 en el segundo.

Si el experimento de lanzar los dos dados se repite muchas veces, la frecuencia relativa del suceso "*obtener suma* 10" será el número de veces que se obtiene el resultado 46 más el número de veces que se obtiene 55 más el número de veces que se obtiene 64, dividido por el número de veces que se realiza el experimento, es decir,

$$f_A = \frac{n_{46} + n_{55} + n_{64}}{n} = \frac{n_{46}}{n} + \frac{n_{55}}{n} + \frac{n_{64}}{n}.$$

Ahora bien, como n_{46}/n es la frecuencia relativa del resultado 46, y lo mismo ocurre para los otros dos resultados, se obtiene que:

$$f_A = f_{46} + f_{55} + f_{64}.$$

Como, además, las probabilidades son las frecuencias relativas cuando el experimento se repite un gran número de veces y si se denota por $p(A)$ a la probabilidad de que ocurra el suceso A, resulta que:

$$p(A) = p(46) + p(55) + p(64).$$

En definitiva, se puede afirmar que la probabilidad de un suceso es la suma de las probabilidades de los resultados que lo componen. Obsérvese que se utiliza la palabra **resultados** para designar a los elementos del espacio muestral y la palabra **suceso** para designar un conjunto de resultados. Como se ha dicho si el suceso consta de un solo resultado se llama elemental.

En el caso de un espacio muestral equiprobable, si el espacio muestral tiene n resultados, cada uno de ellos tiene probabilidad $1/n$. En el ejemplo anterior:

$$p(A) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36} = \frac{1}{12}.$$

Así pues, la probabilidad de un suceso en el caso de un espacio equiprobable es igual al número de elementos del suceso (3 en el ejemplo) dividido por el número de elementos del espacio muestral (36 en el ejemplo). Esta regla fue formulada por el matemático francés Pierre Simon de Laplace diciendo que "la probabilidad de un suceso es igual al número de resultados favorables dividido por el número de resultados posibles":

$$p = \frac{\text{n}^{\text{o}} \text{ de resultados favorables}}{\text{n}^{\text{o}} \text{ de resultados posibles}} \qquad \text{(Regla de Laplace).}$$

Esta fórmula da la probabilidad de cualquier suceso cuando se ha conseguido formular el espacio muestral como un conjunto de resultados equiprobables.

El suceso que contiene todos los elementos del espacio muestral, es decir, el mismo espacio muestral considerado como suceso, se llama **suceso seguro**. El suceso que no contiene ninguno de los resultados posibles se llama **suceso imposible**. La probabilidad del suceso seguro es igual a 1 y la del suceso imposible es igual a 0.

Ejemplo 3 Calcular la probabilidad de sacar suma 7 en el lanzamiento de 2 dados. El espacio muestral consta de 36 resultados y es equiprobable:

$$\begin{aligned} E =&\{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34, 35, 36, \\ &41, 42, 43, 44, 45, 46, 51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}. \end{aligned}$$

El suceso "*sacar suma 7*" es: $S = \{16, 25, 34, 43, 52, 61\}$ y está formado por 6 resultados. Según la regla de Laplace, su probabilidad es:

$$p(S) = \frac{6}{36} = \frac{1}{6}.$$

Ejemplo 4 Calcular la probabilidad de que al extraer al azar dos cartas de una baraja española, se obtengan 2 reyes.

El espacio muestral está formado por todos los posibles modos de sacar dos cartas, o sea $C_{40,2} = \binom{40}{2}$ maneras equiprobables. El suceso "*sacar 2 reyes*" está compuesto por $C_{4,2} = \binom{4}{2}$ resultados. Aplicando la regla de Laplace se tiene:

$$p(\text{“}\textit{sacar 2 reyes}\text{”}) = \frac{\binom{4}{2}}{\binom{40}{2}} = \frac{\frac{4\cdot 3}{2}}{\frac{40\cdot 39}{2}} = \frac{4\cdot 3}{40\cdot 39} = \frac{1}{130}.$$

4. Operaciones con sucesos

A partir de unos sucesos pueden formarse otros. A continuación definiremos el suceso contrario de uno dado, así como los sucesos unión e intersección, $A \cup B$ y $A \cap B$, que se obtienen a partir de dos sucesos A y B.

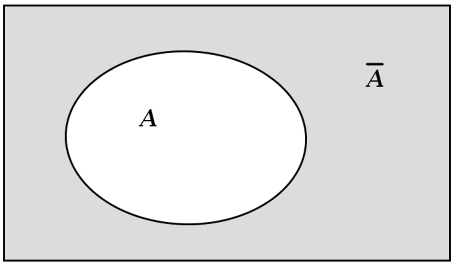

Figura 1: Sucesos A y $\bar{A}$

El conjunto de resultados que no forman parte del suceso A se llama **suceso contrario** de A y se representa por $\bar{A}$. Por ejemplo, en el experimento de lanzar un dado y considerar el suceso "*sacar número mayor o igual que* 5", es decir, $A = \{5, 6\}$; su suceso contrario es "*sacar número menor que* 5", esto es, $\bar{A} = \{1, 2, 3, 4\}$.

Como la suma de las probabilidades de todos los resultados del espacio muestral es igual a 1, la suma de las probabilidades $p(A)$ y $p(\bar{A})$ será igual a 1, puesto que todos los resultados están bien en A o bien en $\bar{A}$. De aquí se deduce que:

$$p(\bar{A}) = 1 - p(A).$$

Para entender en qué consiste el suceso contrario, así como las operaciones con sucesos, conviene representar los sucesos en un diagrama. El espacio muestral se representa como un rectángulo y cada suceso como una superficie dentro del rectángulo. En la Figura 1 la parte sombreada representa el suceso contrario de A. En estos diagramas, debe entenderse que los resultados de un suceso son los puntos de la superficie que lo define, y que la probabilidad del suceso es el cociente entre el área de esa superficie y la de rectángulo que representa el espacio muestral. Si se toma ésta última como unidad, la probabilidad de un suceso es simplemente su área.

Dados dos sucesos A y B, se llama **suceso unión**, $A \cup B$ (se lee "A ó B",) al suceso que contiene todos los resultados que están contenidos en alguno de los dos sucesos, es decir, el suceso $A \cup B$ se cumple si se cumplen A, B o ambos. En la Figura 2 se ha representado sombreado el suceso $A \cup B$. Por ejemplo, en

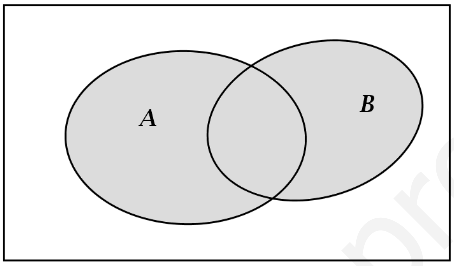

Figura 2: Suceso $A \cup B$

el lanzamiento de dos dados, sean:

$$\begin{aligned} A &= \text{"}\textit{obtener suma 10}\text{"} = \{46, 55, 64\}. \\ B &= \text{"}\textit{obtener dos números iguales}\text{"} = \{11, 22, 33, 44, 55, 66\}. \end{aligned}$$

El suceso $A \cup B$ sería:

$$A \cup B = \{46, 55, 64, 11, 22, 33, 44, 66\}.$$

El **suceso intersección**, $A \cap B$ (se lee "A y B"), está compuesto por los resultados que forman parte de los dos sucesos de forma simultánea. En el ejemplo anterior:

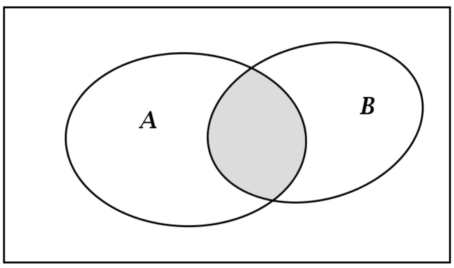

Figura 3: Suceso $A \cap B$

$$\begin{aligned} A &= \text{"}\textit{obtener suma 10}\text{"} = \{46, 55, 64\}. \\ B &= \text{"}\textit{obtener dos números iguales}\text{"} = \{11, 22, 33, 44, 55, 66\}. \\ A \cap B &= \{55\}. \end{aligned}$$

Recordando que en estos diagramas la probabilidad de un suceso se representa como el área de la superficie que representa el suceso cuando se toma como unidad el área del rectángulo, se tiene la siguiente relación entre las probabilidades (ver las Figuras 2 y 3):

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$

En la fórmula anterior hay que restar $p(A \cap B)$ porque si se suman $p(A)$ y $p(B)$, el área correspondiente a $A \cap B$ se contabiliza 2 veces.

Si los sucesos A y B no tienen resultados comunes, se llaman **incompatibles** (no pueden cumplirse en el mismo experimento). En este caso, $p(A \cap B) = 0$ y se tiene que:

$$p(A \cup B) = p(A) + p(B) \qquad (A, B \text{ incompatibles}).$$

El suceso **diferencia** $A - B$ se produce cuando se realiza A y no se realiza B. Equivale a $A \cap \bar{B}$.

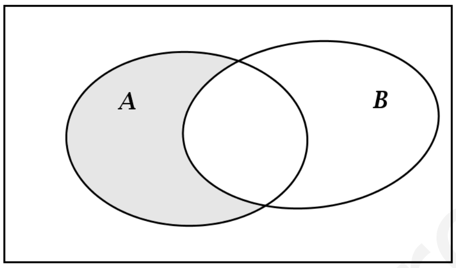

Figura 4: Suceso $A - B$

De la figura 4 se deduce que la probabilidad de $A - B$ es:

$$p(A) = p(A \cap B) + p(A - B) \quad \Longrightarrow \quad p(A - B) = p(A) - p(A \cap B)$$

5. Probabilidad condicionada. Teorema de Bayes

La probabilidad de que se cumplan A y B o sea, la probabilidad de $A \cap B$ puede considerarse compuesta de la probabilidad de que suceda A y de la probabilidad de que se cumpla B una vez que haya sucedido A.

Por ejemplo, puede entenderse que el suceso "*sacar* 2 *ases*" en la extracción de 2 cartas de una baraja de 40 cartas, como 2 extracciones sucesivas de modo que debe salir as en la primera extracción y, una vez obtenido el primer as, debe extraerse el segundo as entre las 39 cartas restantes. La probabilidad de $A \cap B$ se puede obtener a partir de la probabilidad de A y de la de B una vez que ya ha sucedido A. A esta probabilidad de que suceda B una vez que haya sucedido A se le llama probabilidad de B **condicionada** a A y se representa por $p(B/A)$.

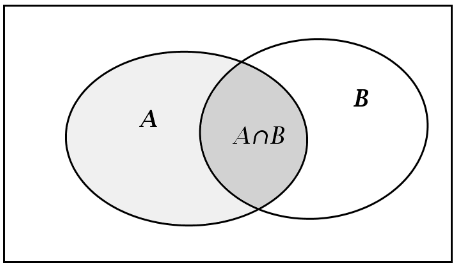

Figura 5: Probabilidad condicionada

Si se sabe que ha sucedido A y se desea saber la probabilidad de B, los únicos resultados posibles son los resultados del suceso A, y los únicos resultados de B que pueden darse son los resultados comunes con A, es decir, los resultados de $A \cap B$. Es como si el espacio muestral estuviese formado por los resultados de A (parte sombreada en la Figura 5) y el suceso B lo formasen los resultados de $A \cap B$ (sombreado mas oscuro). Entonces, comparando las áreas, se tiene que:

$$p(B/A) = \frac{\text{área de } A \cap B}{\text{área de } A} = \frac{p(A \cap B)}{p(A)}.$$

De forma que para la probabilidad de $A \cap B$ se tiene:

$$p(A \cap B) = p(A)p(B/A).$$

Si la probabilidad de B no depende de que A haya sucedido o no, es decir, si $p(B/A) = p(B)$, los sucesos A y B se llaman **independientes** y en ese caso se verifica que:

$$p(A \cap B) = p(A)p(B) \qquad (A, B \text{ independientes}).$$

La regla para calcular la probabilidad de $A \cap B$ por medio de la probabilidad condicionada permite a menudo simplificar los cálculos, tratando el problema como pruebas sucesivas en un espacio muestral más sencillo.

Ejemplo 5 Calcular la probabilidad de que al extraer 3 cartas de una baraja de 40 cartas, resulten las tres de espadas.

Método 1. Considérese el espacio muestral de todas las combinaciones posibles de 3 cartas. Este espacio es equiprobable y está compuesto de $\binom{40}{3}$ resultados posibles. Los casos favorables son las combinaciones de 3 elementos que se pueden dar con las 10 cartas de espadas, de forma que:

$$p = \frac{\binom{10}{3}}{\binom{40}{3}} = \frac{\frac{10\cdot 9\cdot 8}{3!}}{\frac{40\cdot 39\cdot 38}{3!}} = \frac{10 \cdot 9 \cdot 8}{40 \cdot 39 \cdot 38} = \frac{3}{247}.$$

Método 2. Si se consideran 3 extracciones sucesivas, se multiplica la probabilidad de que en la primera extracción resulte una espada, por la probabilidad de que en la segunda extracción resulte otra espada, supuesto que ha salido una en la primera (por tanto quedan 9 espadas entre las 39 cartas), por la probabilidad de que resulte otra espada en la tercera extracción supuesto que han salido 2 espadas en las dos primeras extracciones:

$$p = \frac{10}{40} \cdot \frac{9}{39} \cdot \frac{8}{38} = \frac{3}{247}.$$

Hay que notar que en la primera extracción el espacio muestral tenía 40 resultados, en la segunda 39 y en la tercera 38.

En muchos problemas es preciso aplicar la regla de la suma y el producto de probabilidades sucesivamente. Supóngase que un suceso puede ocurrir de dos formas diferentes incompatibles entre sí. Por ejemplo, una determinada prueba para detectar una enfermedad puede dar un resultado positivo con una determinada probabilidad si el paciente padece la enfermedad, pero también puede dar positivo, con otra probabilidad si el individuo está sano.

Si A_1 es el suceso "*padecer la enfermedad*", A_2 es "*no padecer la enfermedad*" y S es "*obtener un resultado positivo en la prueba*", entonces, el esquema del problema sería el que se muestra en la Figura 6. Como $A_1 \cap S$ y $A_2 \cap S$ son incompatibles, es $p(S) = p(A_1 \cap S) + p(A_2 \cap S)$, y aplicando ahora la regla del producto de probabilidades, se tiene:

$$p(S) = p(A_1 \cap S) + p(A_2 \cap S) = p(A_1)p(S/A_1) + p(A_2)p(S/A_2).$$

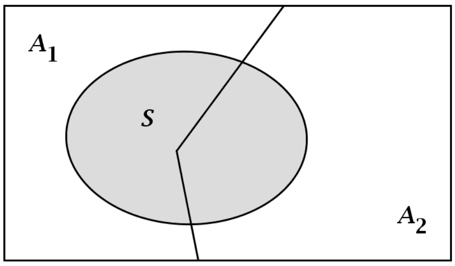

Figura 6: Probabilidad total

Ejemplo 6 En una población, el 10 % de sus habitantes padece una determinada enfermedad. Existe una prueba para detectar la enfermedad que da resultado positivo cuando se aplica a enfermos en el 95 % de los casos. La prueba también da positivo si se aplica a individuos sanos el 10 % de las veces. Calcular la probabilidad de que al aplicar la prueba a un individuo de la población elegido al azar, de resultado positivo.

Sean S = "Obtener resultado positivo al aplicar la prueba", A_1 = "La persona elegida al azar está enferma" y A_2 = "La persona elegida al azar está sana". Aplicando la fórmula anterior resulta que

$$p(S) = p(A_1)p(S/A_1) + p(A_2)p(S/A_2) = 0{,}10 \cdot 0{,}95 + 0{,}90 \cdot 0{,}10 = 0{,}185.$$

En situaciones como la del ejemplo anterior, lo que se pretende es, una vez pasada la prueba y habiendo obtenido un resultado positivo, saber cuál es la probabilidad de que el individuo esté enfermo, es decir, la probabilidad de A_1 cuando ha sucedido S, o sea, $p(A_1/S)$.

Si ha sucedido S, hay que considerar a S como espacio muestral (parte sombreada en la Figura 6). Teniendo en cuenta cómo se obtenía la probabilidad condicionada, resulta lo que se conoce como Teorema de Bayes:

$$p(A_1/S) = \frac{p(A_1 \cap S)}{p(S)} = \frac{p(A_1)p(S/A_1)}{p(A_1)p(S/A_1) + p(A_2)p(S/A_2)}.$$

Cuando un resultado se puede obtener de diversas formas incompatibles (Laplace utilizaba la palabra causas), una vez producido el resultado, el Teorema de Bayes da la probabilidad de que se haya producido a través de una forma determinada.

Ejemplo 7 En el ejemplo anterior, si se supone que la prueba ha dado un resultado positivo, ¿cuál es la probabilidad de que el individuo padezca la enfermedad?

Con la notación del ejemplo anterior, resulta que:

$$p(A_1/S) = \frac{p(A_1)p(S/A_1)}{p(A_1)p(S/A_1) + p(A_2)p(S/A_2)} = \frac{0{,}10 \cdot 0{,}95}{0{,}10 \cdot 0{,}95 + 0{,}90 \cdot 0{,}10} \approx 0{,}514.$$

Apéndice

Paradoja del cumpleaños

Una persona cualquiera, cada vez que toma una decisión resuelve de alguna manera un problema de probabilidad aunque desconozca el contenido matemático del término. La mente humana está preparada para examinar los casos favorables y desfavorables intuitivamente, sin necesidad de realizar operaciones matemáticas y decidir en consecuencia. Cuando la solución matemática de un problema es contraria a este conocimiento intuitivo es cuando puede hablarse de paradoja y una de estas paradojas es el problema del cumpleaños que se va a tratar a continuación.

En el problema del cumpleaños se trata de calcular el número de personas que deben reunirse para que la probabilidad de que dos de ellas cumplan años el mismo día sea al menos del 50 %. Para simplificar el problema, se supone que los nacimientos se distribuyen uniformemente a lo largo del año de forma que la probabilidad de que una persona elegida al azar cumpla años en un día concreto es de 1/365. Se invita al lector a que antes de proseguir la lectura formule una hipótesis sobre la solución del problema. ¿Cuántas personas deben reunirse?, ¿40?, ¿80?, ¿150?

Supóngase que se reúnen 5 personas A_1, A_2, A_3, A_4 y A_5. Para calcular la probabilidad de que 2 de ellas celebren el cumpleaños el mismo día lo más sencillo es calcular previamente la probabilidad del suceso contrario, es decir, la probabilidad de los cinco cumplan años en días diferentes. La probabilidad de que A_2 celebre su cumpleaños en un día diferente que A_1 es de 364/365, la probabilidad de que A_3 lo haga en un día distinto que A_1 y A_2 es de 363/365 y razonando de forma similar para A_4 y A_5 se obtiene, aplicando la regla del producto de probabilidades, que

$$\frac{364}{365}\cdot\frac{363}{365}\cdot\frac{362}{365}\cdot\frac{361}{365}$$

es la probabilidad de que las 5 personas cumplan años en días diferentes. La probabilidad $p(5)$ de que haya al menos 2 de ellas que cumplan años el mismo día es:

$$p(5) = 1 - \frac{364}{365}\cdot\frac{363}{365}\cdot\frac{362}{365}\cdot\frac{361}{365}.$$

Para obtener un valor aproximado de este número conviene escribirlo de la siguiente forma:

$$p(5) = 1 - \left(1-\frac{1}{365}\right)\cdot\left(1-\frac{2}{365}\right)\cdot\left(1-\frac{3}{365}\right)\cdot\left(1-\frac{4}{365}\right).$$

En general, si se reúnen n personas, la probabilidad de que haya al menos una coincidencia en los cumpleaños es:

$$p(n) = 1 - \left(1-\frac{1}{365}\right)\cdot\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{n-1}{365}\right).$$

Para ver cuándo la probabilidad alcanza el valor del 50 %, basta dar valores a n. Los resultados se muestran en la Tabla 1.

n	$p(n)$	n	$p(n)$	n	$p(n)$	n	$p(n)$	n	$p(n)$
2	0.0027	11	0.1411	20	0.4114	29	0.6810	38	0.8641
3	0.0082	12	0.1670	21	0.4437	30	0.7063	39	0.8782
4	0.0164	13	0.1944	22	0.4757	31	0.7305	40	0.8912
5	0.0271	14	0.2231	**23**	**0.5073**	32	0.7533	50	0.9704
6	0.0405	15	0.2529	24	0.5383	33	0.7750	57	0.9901
7	0.0562	16	0.2836	25	0.5687	34	0.7953	70	0.999160
8	0.0743	17	0.3150	26	0.5982	35	0.8144	80	0.999914
9	0.0946	18	0.3469	27	0.6269	36	0.8322	90	0.999994
10	0.1169	19	0.3791	28	0.6545	37	0.8487	100	0.999999

Cuadro 1: Probabilidad de coincidencia de cumpleaños para n personas

Se observa que a partir de $n = 23$, la probabilidad de que coincidan dos cumpleaños es mayor del 50 %. A partir de $n = 57$ la probabilidad es mayor de 99 %. Estos son los valores que, según parece, son contrarios a la intuición y que hacen que el problema merezca el nombre de *paradoja*.

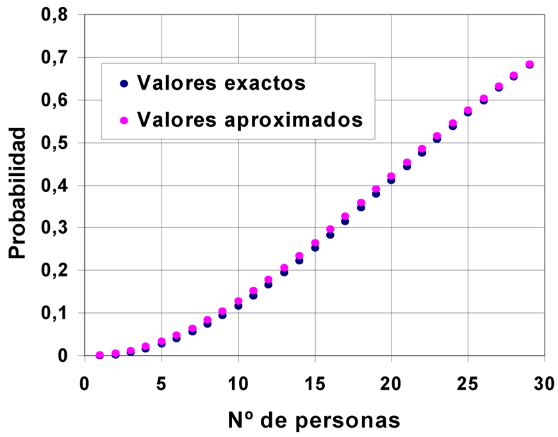

Figura 7: Valores exactos y aproximados de la probabilidad

Resulta interesante obtener una expresión aproximada para $p(n)$ de forma que pueda ser utilizada en otras situaciones. Para ello se usará la aproximación

$$e^x \approx 1 + x.$$

válida cuando x es un número próximo a cero. Esta aproximación se puede deducir de la definición del número e como límite. Habitualmente se define:

$$e = \lim_{t\to\infty} \left(1 + \frac{1}{t}\right)^t.$$

Llamando $x = \frac{1}{t}$, si $t \to \infty$ entonces $x \to 0$, de modo que el número e puede definirse también como:

$$e = \lim_{x\to 0} (1+x)^{\frac{1}{x}}.$$

Esto quiere decir que si x es un número muy pequeño $e \approx (1+x)^{\frac{1}{x}}$ y elevando ambos miembros a x resulta la aproximación buscada $e^x \approx 1 + x$.

Aplicando esta aproximación se tiene que:

$$1 - \frac{1}{365} \approx e^{-\frac{1}{365}}, \quad 1 - \frac{2}{365} \approx e^{-\frac{2}{365}}, \quad \ldots, \quad 1 - \frac{n-1}{365} \approx e^{-\frac{n-1}{365}},$$

de forma que, sustituyendo en la fórmula de $p(n)$ se obtiene:

$$\begin{aligned} p(n) &\approx 1 - e^{-\frac{1}{365}} e^{-\frac{2}{365}} \ldots e^{-\frac{n-1}{365}} \\ &= 1 - e^{-\frac{1}{365} - \frac{2}{365} - \cdots - \frac{n-1}{365}} \\ &= 1 - e^{-\frac{1}{365}(1+2+\ldots+(n-1))} \\ &= 1 - e^{-\frac{1}{365}\frac{n(n-1)}{2}}. \end{aligned}$$

Si n es pequeño frente a n^2 la fórmula anterior puede escribirse de una forma todavía más sencilla como:

$$p(n) = 1 - e^{-\frac{n^2}{2\cdot 365}}.$$

Esta aproximación es bastante buena como puede verse en la Figura 7.

En general, si un experimento aleatorio puede tener k resultados igualmente probables, la probabilidad de que al repetir el experimento n veces se obtenga algún resultado repetido viene dada aproximadamente por la fórmula:

$$p(n) = 1 - e^{-\frac{n^2}{2k}}.$$

Para que la probabilidad sea del 50 % el número de veces que debe repetirse el experimento se obtiene sustituyendo $p(n)$ en la fórmula anterior por $1/2$. Despejando n resulta:

$$n = \sqrt{2k \ln 2} \qquad (p(n) = 1/2),$$

es decir, si el número de resultados posibles k es grande, el número de veces que hay que repetir el experimento para que la probabilidad de obtener algún resultado repetido sea del 50 % es del orden de $\sqrt{k}$.

Ejemplo 8 ¿Cuál es la probabilidad de que se repita la combinación ganadora en 1000 sorteos de la lotería primitiva? ¿Cuántos sorteos deben celebrarse para que la probabilidad de se repita 2 veces la misma combinación sea del 50 %?

La lotería primitiva consiste en acertar una combinación de 6 números entre todas las que pueden formarse con los números comprendidos entre 1 y 49. El número de estas combinaciones es

$$C_{49,6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13983816.$$

(Desde luego que el hecho de que se juegue a la lotería primitiva conociendo este número no deja de ser otra paradoja). Hay $k = 13983816$ resultados igualmente probables, de forma que la probabilidad de que se repita alguno de ellos en $n = 1000$ sorteos está dada por

$$p(1000) \approx 1 - e^{-\frac{1000^2}{2 \cdot 13983816}} = 0,0351.$$

La probabilidad es entonces del 3,51 % aproximadamente.

Para que la probabilidad de que se obtenga la misma combinación ganadora dos veces sea del 50 %, el número de sorteos que deberán celebrarse es

$$n = \sqrt{2 \cdot 13983816 \cdot \ln 2} = 4402,92,$$

es decir, unos 4403 sorteos. Se celebran dos sorteos semanales de modo que, en promedio, se repite una combinación ganadora cada 42 años.