

ESPACIOS DE PROBABILIDAD

Julián de la Horra

Departamento de Matemáticas U.A.M.

1 Introducción

Los espacios de probabilidad son las estructuras que se utilizan en Matemáticas para estudiar los fenómenos aleatorios, es decir, aquellos fenómenos que tienen un alto grado de incertidumbre en sus resultados.

Consideremos, por ejemplo, una situación muy sencilla: el lanzamiento de una moneda. Sería posible, quizá, intentar predecir el resultado del lanzamiento (cara o cruz) desarrollando un modelo muy complejo que tuviera en cuenta variables como el ángulo de lanzamiento, la fuerza con que se lanza, el rozamiento del aire, el material de la moneda, el material de la superficie en la que va a caer,... Pero, aparte de la enorme complejidad que tendría un modelo de este tipo, en general, nos va a resultar suficiente para muchas cosas, conocer la “probabilidad” del resultado.

Los modelos de probabilidad trabajan con las “probabilidades” de los diferentes resultados o sucesos que nos puedan interesar en un fenómeno aleatorio, y nos dan respuestas más que satisfactorias en el estudio de dichos fenómenos.

Una observación inicial: al desarrollar la teoría de la Probabilidad y explicar sus resultados, es muy frecuente recurrir a ejemplos como el lanzamiento de monedas o de dados, la extracción de cartas de una baraja,... No se trata de que esos sean los problemas interesantes que nos resuelve la teoría de la Probabilidad (que no lo son, salvo que uno sea un tahúr), sino que son ejemplos sencillos de plantear y de visualizar y, por ese motivo, ayudan a comprender lo que se está haciendo. El enorme interés de la Teoría de la Probabilidad reside en que, aparte de tener (en mi opinión) unos preciosos desarrollos matemáticos, permite estudiar, a través de la Inferencia Estadística, los problemas que surgen en los estudios científicos que intentan obtener conclusiones a partir de datos reales.

2 Espacio muestral y σ -álgebra de sucesos

Definición.- El *espacio muestral*, Ω , es el conjunto de todos los posibles resultados del fenómeno aleatorio que deseamos estudiar. •

Definición.- Una σ -álgebra de sucesos, S , es un subconjunto de partes de Ω que verifica:

- (a) Si $A \in S$, entonces $A^c \in S$.
- (b) Si $A_1, \dots, A_n, \dots \in S$, entonces $\bigcup_n A_n \in S$. •

La idea de una σ -álgebra es que contenga todos los sucesos en los que podemos estar interesados en un fenómeno aleatorio. Vamos a ver a continuación, que en una σ -álgebra están todos los conjuntos que se obtienen a partir de las operaciones corrientes realizadas con conjuntos.

Propiedades.- Para cualquier σ -álgebra S , tenemos:

- (a) $\emptyset \in S, \Omega \in S$.
- (b) Si $A_1, \dots, A_n, \dots \in S$, entonces $\bigcap_n A_n \in S$.
- (c) Si $A_1, A_2 \in S$, entonces $A_1 - A_2 \in S$.
- (d) Si $A_1, \dots, A_n, \dots \in S$, entonces $\overline{\lim} A_n = \bigcap_{n=1}^{\infty} [\bigcup_{m=n}^{\infty} A_m] \in S$ y $\underline{\lim} A_n = \bigcup_{n=1}^{\infty} [\bigcap_{m=n}^{\infty} A_m] \in S$.
- (e) Si $A_1, \dots, A_n, \dots \in S$, entonces $\lim A_n \in S$ (cuando este límite existe).
- (f) La intersección de una familia de σ -álgebras sigue siendo una σ -álgebra.

Demostración.-

- (a) Tomamos $A \in S \Rightarrow A^c \in S \Rightarrow \Omega = A \cup A^c \in S \Rightarrow \emptyset = \Omega^c \in S$.
- (b) $\bigcap_n A_n = [\bigcup_n A_n^c]^c \in S$.
- (c) $A_1 - A_2 = A_1 \cap A_2^c \in S$.
- (d) Inmediato.
- (e) Inmediato ya que, si $\lim A_n$ existe: $\lim A_n = \overline{\lim} A_n = \underline{\lim} A_n$.
- (f) Inmediato. •

Ejemplos.- Las σ -álgebras más utilizadas en la práctica son las siguientes:

(a) Cuando el espacio muestral, Ω , es finito o numerable, la σ -álgebra habitualmente utilizada es partes de Ω . Es decir, consideramos que cualquier subconjunto de Ω puede ser un suceso de interés. Esta elección es la más sencilla y no ofrece ningún problema.

(b) Cuando el espacio muestral es $\Omega = \mathbb{R}$, no se utiliza partes de Ω , porque esta σ -álgebra no es adecuada para construir una función de probabilidad. En este caso, la σ -álgebra habitualmente utilizada es la σ -álgebra de Borel que es la intersección de todas las σ -álgebras que contienen a todos los intervalos de la forma $(-\infty, x]$ o, dicho de otra manera, la mínima σ -álgebra que contiene a todos los intervalos de la forma $(-\infty, x]$. Se suele simplificar diciendo que la

σ -álgebra de Borel está generada por la clase $C = \{(-\infty, x]\}$. Es inmediato darse cuenta de que cualquier otro intervalo también estará en la σ -álgebra de Borel.

(c) Cuando el espacio muestral es $\Omega = \mathbb{R}^n$, tampoco utilizamos partes de Ω , porque esta σ -álgebra tampoco es adecuada para construir una función de probabilidad. En este caso, la σ -álgebra habitualmente utilizada es la σ -álgebra de Borel en \mathbb{R}^n que es la σ -álgebra generada por la clase $C = \{(-\infty, x_1] \times \dots \times (-\infty, x_n]\}$.

3 Modelo de probabilidad

Con todo lo anterior tenemos ya el sustrato necesario para poder definir un modelo de probabilidad.

Definición.- Consideramos un espacio muestral Ω y una σ -álgebra S . Un *modelo de probabilidad* o función de probabilidad o medida de probabilidad es una función $P : S \rightarrow [0, 1]$, que verifica:

- (a) $P(\Omega) = 1$
- (b) Si $A_1, \dots, A_n, \dots \in S$, y son disjuntos, entonces:

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n) \quad \bullet$$

La manera más sencilla de entender el motivo de esta definición es pensar que el concepto de probabilidad trata de formalizar lo que ocurre con las frecuencias relativas de los distintos sucesos que podemos encontrarnos al lanzar un dado con 6 caras numeradas del 1 al 6:

- La frecuencia relativa de cualquier suceso siempre es un número entre 0 y 1.
- La frecuencia relativa de que salga algún número de los posibles (es decir, entre 1 y 6) es, evidentemente, 1.
- La frecuencia relativa de la unión de sucesos disjuntos es la suma de sus frecuencias relativas.

Ahora bien, la comparación que se acaba de hacer con las frecuencias relativas no significa que esa sea la única interpretación posible del concepto de probabilidad; es, simplemente, una de las interpretaciones más sencillas de explicar y de visualizar.

La terna (Ω, S, P) suele recibir el nombre de **espacio de probabilidad**, y será un caso particular de los espacios de medida que se consideran en la Teoría de la Medida.

Las siguientes propiedades serán muy útiles para llevar a cabo el cálculo de probabilidades en diferentes situaciones.

Propiedades.- Para cualquier modelo de probabilidad P , tenemos:

- (a) Si $A \in S$, entonces $P(A^c) = 1 - P(A)$.
- (b) $P(\emptyset) = 0$.
- (c) Si $A_1, A_2 \in S$ y $A_1 \subset A_2$, entonces $P(A_1) \leq P(A_2)$ y $P(A_2 - A_1) = P(A_2) - P(A_1)$.
- (d) Si $A_1, A_2 \in S$, entonces $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
- (e) Si $A_1, \dots, A_n \in S$, entonces $P(\bigcup A_i) = \sum_i P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots + (-1)^{n+1} P(\bigcap_{i=1}^n A_i)$.
- (f) Si $A_1, \dots, A_n \in S$, entonces $P(\bigcup A_i) \leq \sum_i P(A_i)$.

Demostración.-

- (a) Tenemos $A \cup A^c = \Omega \Rightarrow P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1 \Rightarrow P(A^c) = 1 - P(A)$.
- (b) $P(\emptyset) = 1 - P(\Omega) = 1 - 1 = 0$.
- (c) Si $A_1 \subset A_2$, entonces $A_2 = A_1 \cup (A_2 - A_1)$ (disjuntos) $\Rightarrow P(A_2) = P(A_1) + P(A_2 - A_1)$, y los dos resultados son inmediatos.
- (d) Tenemos que $A_1 \cup A_2 = A_1 \cup [A_2 - (A_1 \cap A_2)]$ (disjuntos) $\Rightarrow P(A_1 \cup A_2) = P(A_1) + P[A_2 - (A_1 \cap A_2)] = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
- (e) Se prueba por inducción.
- (f) Se prueba por inducción.

¿Se puede permutar el límite con la probabilidad? Los siguientes resultados contestan a esta pregunta.

Teorema.-

- (a) Sea A_1, \dots, A_n, \dots una sucesión creciente de conjuntos de S . Entonces:

$$P(\lim_n A_n) = \lim_n P(A_n).$$

- (b) Sea A_1, \dots, A_n, \dots una sucesión decreciente de conjuntos de S . Entonces:

$$P(\lim_n A_n) = \lim_n P(A_n).$$

Demostración.-

(a) Para una sucesión creciente de conjuntos: $\lim_n A_n = \bigcup_{n=1}^{\infty} A_n$.

Consideramos la siguiente sucesión de conjuntos disjuntos:

$$B_1 = A_1, B_2 = A_2 - A_1, \dots, B_n = A_n - A_{n-1}, \dots$$

Tenemos:

$$\begin{aligned} P(\lim_n A_n) &= P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) = \lim_n [\sum_{i=1}^n P(B_i)] \\ &= \lim_n P(\bigcup_{i=1}^n B_i) = \lim_n P(A_n). \end{aligned}$$

(b) Para una sucesión decreciente de conjuntos: $\lim_n A_n = \bigcap_{n=1}^{\infty} A_n$.

Consideramos la siguiente sucesión creciente de conjuntos:

$$B_1 = A_1 - A_2, B_2 = A_1 - A_3, \dots, B_n = A_1 - A_{n+1}, \dots$$

y tenemos que $\lim_n B_n = A_1 - \lim_n A_n$. Entonces:

$$\begin{aligned} P(A_1) - P(\lim_n A_n) &= P(A_1 - \lim_n A_n) = P(\lim_n B_n) = \lim_n P(B_n) \\ &= \lim_n P(A_1 - A_{n+1}) = \lim_n [P(A_1) - P(A_{n+1})] \\ &= P(A_1) - \lim_n P(A_{n+1}). \end{aligned}$$

Por tanto, tenemos que $P(\lim_n A_n) = \lim_n P(A_n)$. •

Teorema.- Sea A_1, \dots, A_n, \dots una sucesión de conjuntos de S . Entonces:

$$P(\underline{\lim} A_n) \leq \underline{\lim} P(A_n) \leq \overline{\lim} P(A_n) \leq P(\overline{\lim} A_n).$$

Demostración.- Vamos a demostrar la primera desigualdad, ya que la tercera tiene una demostración análoga y la segunda es una obviedad:

$$P(\underline{\lim} A_n) = P(\bigcup_{n=1}^{\infty} [\bigcap_{m=n}^{\infty} A_m]) = P(\lim_n [\bigcap_{m=n}^{\infty} A_m]) = \lim_n P(\bigcap_{m=n}^{\infty} A_m) (*)$$

Para todo $m \geq n$: $P(\bigcap_{m=n}^{\infty} A_m) \leq P(A_m)$.

Por tanto: $P(\bigcap_{m=n}^{\infty} A_m) \leq \inf_{m \geq n} P(A_m)$.

$$(*) \leq \lim_n [\inf_{m \geq n} P(A_m)] = \underline{\lim} P(A_n). \quad \bullet$$

Corolario.- Sea A_1, \dots, A_n, \dots una sucesión de conjuntos de S . Si la sucesión tiene límite, entonces:

$$P(\lim_n A_n) = \lim_n P(A_n).$$

Demostración.- Si la sucesión tiene límite, tenemos:

$$\lim_n A_n = \underline{\lim} A_n = \overline{\lim} A_n$$

El resultado es ahora inmediato, aplicando el teorema anterior. •

Podemos considerar modelos de probabilidad sobre todo tipo de espacios muestrales pero, naturalmente, conviene centrarse en los más habituales. Eso es lo que vamos a hacer en las tres siguientes secciones.

4 Espacio muestral discreto

En esta sección, consideramos que el espacio muestral, Ω , es discreto, es decir, es finito o numerable aunque, por simplicidad, se va a utilizar únicamente la notación correspondiente al caso numerable. Si es finito, bastará con quitar los puntos suspensivos:

$$\Omega = \{a_1, \dots, a_n, \dots\}$$

En estos casos, ya hemos dicho anteriormente que la σ -álgebra habitualmente utilizada es partes de Ω . Es decir, consideramos que cualquier subconjunto de Ω puede ser un suceso de interés.

En este caso, para tener perfectamente definido un modelo de probabilidad, no es necesario detallar de entrada la probabilidad de cualquier suceso, sino que es suficiente con definir la función de masa asociada al modelo de probabilidad:

Definición.- La *función de masa* asociada a un modelo de probabilidad P sobre un espacio muestral (finito o numerable) $\Omega = \{a_1, \dots, a_n, \dots\}$ es la función $P : \Omega \rightarrow [0, 1]$, que verifica:

- (a) $P(a_n) \in [0, 1]$, para todo $a_n \in \Omega$.
- (b) $\sum_n P(a_n) = 1$. •

Es decir, la función de masa nos proporciona la probabilidad de todos los sucesos elementales (sucesos formados por un único elemento). A partir de la función de masa, obtener la probabilidad de cualquier subconjunto A

de Ω (finito o infinito) es inmediato, utilizando las propiedades de cualquier modelo de probabilidad:

$$P(A) = \sum_{a_i \in A} P(a_i).$$

A partir de esta definición, es inmediato comprobar que se verifican las dos propiedades que tiene que cumplir cualquier modelo de probabilidad.

Dentro de los espacios muestrales discretos, tiene especial interés el caso en que Ω es finito, y todos los sucesos elementales (sucesos formados por un único elemento) son considerados equiprobables:

$$\Omega = \{a_1, \dots, a_n\} ; \quad P(a_1) = \dots = P(a_n) = \frac{1}{n}$$

En este caso, para un subconjunto $A = \{a_i, \dots, a_j\}$ con k elementos, tendremos:

$$P(A) = P\{a_i, \dots, a_j\} = P(a_i) + \dots + P(a_j) = \frac{k}{n} = \frac{\text{Casos favorables}}{\text{Casos posibles}} ,$$

que constituye la muy conocida y popular *regla de Laplace*, que sólo se puede aplicar cuando el espacio muestral es finito y los sucesos elementales son equiprobables.

5 Modelos de probabilidad sobre \Re

En esta sección, consideramos la recta real, \Re , como espacio muestral.

Ya hemos dicho anteriormente, que la σ -álgebra habitualmente utilizada sobre la recta real es la σ -álgebra de Borel, que es la mínima σ -álgebra que contiene a todos los intervalos. ¿Cuál es la razón para utilizar esta σ -álgebra en vez de la σ -álgebra de partes de Ω , que podría parecer la más natural? La razón es que para definir un modelo de probabilidad sobre la recta real no necesitamos definir la probabilidad de cada suceso (afortunadamente, ya que sería inviable), sino que es suficiente con definir la probabilidad de ciertos intervalos, es decir, es suficiente con definir la función de distribución asociada al modelo de probabilidad:

Definición.- La *función de distribución* asociada a un modelo de probabilidad P sobre \Re es la función $F : \Re \rightarrow [0, 1]$, definida como:

$$F(x) = P(-\infty, x] \quad \bullet$$

Es decir, la función de distribución nos proporciona las probabilidades de todos los intervalos de la forma $(-\infty, x]$.

En primer lugar, vamos a obtener las propiedades fundamentales que tiene cualquier función de distribución, utilizando las propiedades de los modelos de probabilidad:

Propiedades.- Cualquier función de distribución, F , sobre \Re verifica:

- (a) $\lim_{x \rightarrow -\infty} F(x) = 0$.
- (b) $\lim_{x \rightarrow \infty} F(x) = 1$.
- (c) F es una función creciente.
- (d) F es continua por la derecha.

Demostración.-

- (a) $\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P(-\infty, x] = P(\lim_{x \rightarrow -\infty} (-\infty, x]) = P(\emptyset) = 0$.
- (b) $\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} P(-\infty, x] = P(\lim_{x \rightarrow \infty} (-\infty, x]) = P(\Re) = 1$.
- (c) Para $a < b$: $F(b) = P(-\infty, b] = P((-\infty, a] \cup (a, b]) = P((-\infty, a] + P(a, b]) \geq P((-\infty, a]) = F(a)$.
- (d) $F(x^+) = \lim_{h \rightarrow 0} F(x+h) = \lim_{h \rightarrow 0} P(-\infty, x+h] = P(\lim_{h \rightarrow 0} (-\infty, x+h]) = P(-\infty, x] = F(x)$. •

En segundo lugar, veremos que, a partir de una función de distribución, se puede obtener la probabilidad de cualquier tipo de intervalo, utilizando nuevamente las propiedades de los modelos de probabilidad. Lo vemos con algunos, y los demás son similares:

- (a) $P(-\infty, x) = P(\lim_{h \rightarrow 0} (-\infty, x-h]) = \lim_{h \rightarrow 0} P(-\infty, x-h] = \lim_{h \rightarrow 0} F(x-h) = F(x^-)$.
- (b) $P(a, b] = P((-\infty, b] - (-\infty, a]) = P(-\infty, b] - P(-\infty, a] = F(b) - F(a)$.
- (c) $P(a, b) = P(\lim_{h \rightarrow 0} (a, b-h]) = \lim_{h \rightarrow 0} P(a, b-h] = \lim_{h \rightarrow 0} (F(b-h) - F(a)) = F(b^-) - F(a)$.
- (d) $P(x) = P((-\infty, x] - (-\infty, x)) = P(-\infty, x] - P(-\infty, x) = F(x) - F(x^-)$.

Para obtener estas probabilidades hemos ido aplicando las propiedades que conocemos para cualquier función de probabilidad: la probabilidad de una diferencia de sucesos es la diferencia de probabilidades (cuando uno está contenido en el otro), el límite se puede permutar con la probabilidad en las sucesiones monótonas, ...

Naturalmente, con todo lo anterior sabremos calcular probabilidades de los conjuntos más habituales (por ejemplo, los diferentes tipos de intervalos), pero no de todos los conjuntos que están incluidos en la σ -álgebra de Borel (algunos de ellos, rarísimos). Pero eso no va a importar, ya que la Teoría de la Medida nos ofrecerá el Teorema de Extensión, que nos asegurará que es posible extender la función de probabilidad (de manera única) a la σ -álgebra generada por los intervalos, es decir, a la σ -álgebra de Borel.

6 Modelos de probabilidad sobre \mathfrak{R}^n

En esta sección, consideramos \mathfrak{R}^n como espacio muestral.

Ya hemos dicho anteriormente, que la σ -álgebra habitualmente utilizada en este caso es la σ -álgebra de Borel en \mathfrak{R}^n , que es la mínima σ -álgebra que contiene a todos los rectángulos. Las razones para utilizar esta σ -álgebra son análogas a las esgrimidas para la recta real. En este caso, para tener perfectamente definido un modelo de probabilidad, tampoco es necesario detallar de entrada la probabilidad de cualquier suceso, sino que es suficiente con definir la función de distribución asociada al modelo de probabilidad:

Definición.- La *función de distribución* asociada a un modelo de probabilidad P sobre \mathfrak{R}^n es la función $F : \mathfrak{R}^n \rightarrow [0, 1]$, definida como:

$$F(x_1, \dots, x_n) = P\{(-\infty, x_1] \times \dots \times (-\infty, x_n]\} \quad \bullet$$

Es decir, la función de distribución nos proporciona las probabilidades de todos los rectángulos de la forma $(-\infty, x_1] \times \dots \times (-\infty, x_n]$.

Las siguientes propiedades de una función de distribución en \mathfrak{R}^n son similares a las propiedades de una función de distribución en \mathfrak{R} :

Propiedades.- Cualquier función de distribución, F , sobre \mathfrak{R}^n verifica:

- (a) $\lim F(x_1, \dots, x_i, \dots, x_n) = 0$, cuando alguna x_i tiende a $-\infty$.
- (b) $\lim F(x_1, \dots, x_i, \dots, x_n) = 1$, cuando todas las x_i tienden a ∞ .
- (c) F es una función creciente en cada una de sus variables.
- (d) F es continua por la derecha en cada una de sus variables.

Las demostraciones son similares a las propiedades de las funciones de distribución sobre \mathfrak{R} .

7 Probabilidad condicionada y sucesos independientes

Todos los sucesos que vamos a considerar a continuación tendrán una probabilidad positiva, es decir, excluirémos de las definiciones aquellos sucesos con probabilidad cero.

El concepto de probabilidad de un suceso A condicionada por un suceso B , que vamos a definir a continuación, formaliza la idea de cuál será la probabilidad que tendrá el suceso A , cuando sabemos que ha ocurrido el suceso B , es decir, formaliza la idea intuitiva de que la probabilidad de un suceso siempre depende de la información que tengamos.

Definición.- La *probabilidad de un suceso A condicionada por un suceso B* se define como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \bullet$$

En línea con lo anterior, podemos pensar que hay sucesos que nos dan información sobre otros sucesos, de modo que saber que ha ocurrido B modifica (condiciona) la probabilidad de A . Pero también hay sucesos que no dan ninguna información sobre otros sucesos, de modo que saber que ha ocurrido B no modifica (no condiciona) la probabilidad de A ; en este último caso, los sucesos A y B son intuitivamente independientes uno de otro.

A continuación, se formaliza la definición de sucesos independientes:

Definición.- Los *sucesos A y B decimos que son independientes* cuando:

$$P(A \cap B) = P(A)P(B) \quad \bullet$$

En la siguiente propiedad, relacionamos la definición formal de sucesos independientes con la idea intuitiva de que, si dos sucesos son independientes, las probabilidades condicionadas no deberían verse modificadas.

Propiedad.- Dos sucesos A y B son independientes si y sólo si $P(A|B) = P(A)$ (equivalentemente, $P(B|A) = P(B)$).

Demostración.-

(a) Supongamos que A y B son independientes. Tenemos:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

(b) Supongamos ahora que $P(A|B) = P(A)$. Tenemos:

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B) ,$$

y, por tanto, A y B son independientes. •

Ejemplo.- Vamos a ilustrar las ideas anteriores con un ejemplo sencillo. Consideramos el experimento consistente en dos lanzamientos sucesivos de una moneda equilibrada. El espacio muestral es:

$$\Omega = \{(C, C), (C, Cr), (Cr, C), (Cr, Cr)\}$$

y los cuatro sucesos elementales son equiprobables.

En primer lugar, consideramos los sucesos A = “Cara en el primer lanzamiento” y B = “Cara en el segundo lanzamiento”. Estos dos sucesos son intuitivamente independientes, en el sentido de que saber lo que hemos obtenido en el primer lanzamiento no parece darnos ninguna información relevante sobre el resultado del segundo lanzamiento. Veamos lo que ocurre formalmente:

$$P(A) = P\{(C, C), (C, Cr)\} = 1/4 + 1/4 = 1/2.$$

$$P(B) = P\{(C, C), (Cr, C)\} = 1/4 + 1/4 = 1/2.$$

$$P(A \cap B) = P\{(C, C)\} = 1/4.$$

Por tanto, A y B no sólo son intuitivamente independientes, sino que son formalmente independientes.

Consideramos ahora los sucesos A = “Cara en el segundo lanzamiento” y B = “Exactamente una cara entre los dos lanzamientos”. Si nos informan de que ha ocurrido el suceso B , ¿esta información modifica (condiciona) la probabilidad de A ? Dicho de otra manera, ¿ A y B son sucesos independientes? En este caso, la intuición es menos clara (y puede ser engañosa). Veamos lo que ocurre formalmente:

$$P(A) = P\{(C, C), (Cr, C)\} = 1/4 + 1/4 = 1/2.$$

$$P(B) = P\{(C, Cr), (Cr, C)\} = 1/4 + 1/4 = 1/2.$$

$$P(A \cap B) = P\{(Cr, C)\} = 1/4.$$

Por tanto, en este caso, A y B también son independientes. También podemos calcular directamente la probabilidad condicionada:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P\{(Cr, C)\}}{P\{(C, Cr), (Cr, C)\}} = \frac{1/4}{1/2} = 1/2. \quad \bullet$$

La definición de independencia de sucesos se complica cuando se extiende a más de dos sucesos:

Definición.- Los sucesos $\{A_i\}_{i \in I}$ decimos que son independientes cuando para toda colección finita $\{i_1, \dots, i_k\}$ de índices distintos de I tenemos:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}) \quad \bullet$$

8 Reglas para calcular probabilidades

A continuación, vamos a obtener tres reglas (teoremas sencillos) que resultarán muy útiles para el cálculo de probabilidades en determinadas situaciones.

A partir de la definición de probabilidad condicionada, tenemos que $P(A \cap B) = P(A|B)P(B)$. La siguiente regla generaliza esta expresión a la intersección de n sucesos:

Teorema (regla de la multiplicación).- Consideramos n sucesos, A_1, \dots, A_n . Tenemos:

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P\left(A_n \left| \bigcap_{i=1}^{n-1} A_i \right.\right)$$

Demostración.- La demostración es inmediata: sólo hay que escribir las probabilidades condicionadas del segundo miembro e ir cancelando unos términos con otros. \bullet

Esta regla será muy útil cuando queramos calcular la probabilidad de ocurrencia simultánea de varios sucesos (la probabilidad de la intersección de esos sucesos) y las probabilidades condicionadas del segundo miembro sean sencillas de calcular. Naturalmente, en el caso de que los sucesos A_1, \dots, A_n sean independientes tenemos, simplemente: $P(\bigcap_{i=1}^n A_i) = P(A_1) \dots P(A_n)$.

Ejemplo.- En una urna hay 20 bolas blancas y 10 bolas negras. Hacemos tres extracciones sin devolución a la urna. ¿Cuál es la probabilidad de que las tres sean blancas?

Si llamamos A_i ($i = 1, 2, 3$) al suceso “La i -ésima bola extraída es blanca”, tenemos:

$$\begin{aligned} P(\text{Las tres son blancas}) &= P(A_1 \cap A_2 \cap A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ &= \frac{20}{30} \frac{19}{29} \frac{18}{28} \end{aligned}$$

Las tres probabilidades condicionadas del final son triviales de obtener; por ejemplo, $P(A_3|A_1 \cap A_2)$ es la probabilidad de obtener bola blanca en la tercera extracción, cuando sabemos que las dos primeras han sido blancas; Por tanto, en ese momento, quedan en la urna 28 bolas, de las cuales, 18 son blancas; en consecuencia, $P(A_3|A_1 \cap A_2)=18/28$. •

Teorema (regla de la probabilidad total).- Sean A_1, \dots, A_n sucesos tales que:

(a) $\bigcup_{i=1}^n A_i = \Omega$

(b) $A_i \cap A_j = \emptyset$, para todo $i \neq j$.

Entonces:

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

Demostración.- En efecto:

$$\begin{aligned} P(B) &= P(B \cap (\bigcup_{i=1}^n A_i)) = P(\bigcup_{i=1}^n (B \cap A_i)) = \sum_{i=1}^n P(B \cap A_i) \\ &= \sum_{i=1}^n P(A_i)P(B|A_i) \quad \bullet \end{aligned}$$

Teorema (regla de Bayes).- Sean A_1, \dots, A_n sucesos tales que:

(a) $\bigcup_{i=1}^n A_i = \Omega$

(b) $A_i \cap A_j = \emptyset$, para todo $i \neq j$.

Entonces:

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

Demostración.- En efecto:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)} \quad \bullet$$

La regla de la probabilidad total y la regla de Bayes van a ser especialmente útiles cuando se den las siguientes circunstancias:

(a) El experimento aleatorio se puede separar en dos etapas.

(b) Es sencillo dar una partición de todo el espacio muestral Ω mediante sucesos A_1, \dots, A_n correspondientes a resultados de la primera etapa.

(c) Las probabilidades $P(A_1), \dots, P(A_n)$ son conocidas o fácilmente calculables.

(d) Las probabilidades $P(B|A_1), \dots, P(B|A_n)$ son conocidas o fácilmente calculables, donde B es un suceso correspondiente a resultados de la segunda etapa,

Cuando se den estas circunstancias, la regla de la probabilidad total será muy útil para calcular $P(B)$ y la regla de Bayes será muy conveniente para obtener $P(A_j|B)$.

En los siguientes ejemplos, veremos que, a menudo, el experimento aleatorio considerado corresponde a esta situación.

El primer ejemplo es de urnas y bolas, porque es fácil de comprender.

Ejemplo.- Disponemos de dos urnas. La urna 1 contiene 3 bolas blancas y 2 negras; la urna 2 contiene 2 bolas blancas y 3 negras. Con probabilidad $1/3$, extraeremos una bola al azar de la urna 1 y con probabilidad $2/3$ la extraeremos de la urna 2.

¿Cuál es la probabilidad de que la bola extraída sea blanca?

Si al final nos comunican que la bola extraída ha sido blanca, ¿cuál es la probabilidad de que la extracción se haya efectuado de la urna 1?

Se dan todas las condiciones enumeradas anteriormente:

(a) El experimento consta claramente de dos etapas:

Primera etapa: Elección de la urna.

Segunda etapa: Extracción de la bola.

(b) Se puede hacer, de manera trivial, una partición del espacio muestral mediante los resultados de la primera etapa: A_1 = “La extracción se efectúa de la urna 1” y A_2 = “La extracción se efectúa de la urna 2”.

(c) Tenemos: $P(A_1) = 1/3$ y $P(A_2) = 2/3$.

(d) Si llamamos B al suceso “La bola extraída es blanca”, tenemos: $P(B|A_1) = 3/5$ y $P(B|A_2) = 2/5$.

Por tanto, aplicando la regla de la probabilidad total, tenemos:

$$\begin{aligned} P(\text{Bola extraída es blanca}) &= P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) \\ &= \frac{1}{3} \frac{3}{5} + \frac{2}{3} \frac{2}{5} = \frac{7}{15} \end{aligned}$$

Aplicando ahora la regla de Bayes, la probabilidad de que la extracción se haya efectuado de la urna 1, cuando sabemos que la bola obtenida es blanca, será:

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \\ &= \frac{(1/3)(3/5)}{(1/3)(3/5) + (2/3)(2/5)} = \frac{3}{7} \end{aligned}$$

Según hemos podido ver en el ejemplo, la regla de Bayes nos permite actualizar, de manera automática, la probabilidad sobre diferentes sucesos a la vista de la información adicional que vamos obteniendo con la experimentación. En principio, la probabilidad de que la extracción se efectuase de la urna 1 era $1/3$. Pero una vez que hemos sacado una bola y hemos comprobado que es blanca, parece más probable que estemos utilizando la urna 1 (que contiene más bolas blancas). La regla de Bayes nos permite actualizar esa probabilidad de manera muy sencilla. •

En el siguiente ejemplo se ve todavía mejor cómo la experimentación va modificando nuestras probabilidades.

Ejemplo.- Las plantas de una especie pueden tener flores rojas homocigóticas (RR), rojas heterocigóticas (Rr) o blancas (rr). Aproximadamente, el 70% de las plantas con flores rojas son heterocigóticas. Nos traen una planta con flores rojas y para intentar diagnosticar si es homocigótica o no, la cruzamos con una planta de flores blancas. Si de este cruce obtenemos 5 plantas, todas con flores rojas, ¿cuál es la probabilidad de que fuera homocigótica?

Intuitivamente, si las cinco plantas obtenidas en el cruce son rojas (necesariamente Rr), parece bastante probable que la planta que nos han traído fuera RR ; pero no es seguro. La regla de Bayes nos permite cuantificar esta probabilidad de manera muy sencilla. Observemos primero en qué condiciones nos encontramos:

(a) El experimento consta de dos etapas:

Primera etapa: obtención al azar de la planta que nos traen (en realidad, no realizamos ninguna obtención al azar, pero no hay ningún problema en considerarlo así).

Segunda etapa: Cruce con la planta blanca para obtener 5 nuevas plantas.

(b) Partición del espacio muestral mediante resultados de la primera etapa: RR = “La planta que nos traen es RR ” y Rr = “La planta que nos traen es Rr ”.

(c) Tenemos: $P(RR) = 0,3$ y $P(Rr) = 0,7$.

(d) Si llamamos B al suceso “Las 5 plantas obtenidas en el cruce tienen flores rojas”, tenemos: $P(B|RR) = 1$ y $P(B|Rr) = (1/2)^5$.

Por tanto, la probabilidad pedida es:

$$P(RR|B) = \frac{P(RR)P(B|RR)}{P(RR)P(B|RR) + P(Rr)P(B|Rr)}$$

$$= \frac{(0,3)(1)}{(0,3)(1) + (0,7)(1/2)^5} = 0,932$$

Naturalmente, la probabilidad de ser homocigótica se ha modificado sustancialmente a la vista del resultado obtenido en el experimento: ha pasado de ser 0,3 a valer 0,932. En general, el objetivo de un experimento de este tipo es conseguir la mayor certeza posible, es decir, conseguir que las probabilidades se acerquen a 0 ó a 1. •