

Probabilidad condicional

Consideremos una urna que contiene 4 bolillas rojas y 5 blancas. De las 4 bolillas rojas, 2 son lisas y 2 rayadas y de las 5 bolillas blancas, 4 son lisas y una sola es rayada. Supongamos que se extrae una bolilla y, sin que la hayamos mirado, alguien nos dice que la bolilla es roja, ¿cuál es la probabilidad de que la bolilla sea rayada?

Sean los sucesos A: “la bolilla es rayada” y B: “la bolilla es roja”. Obviamente, sin ninguna información previa, $P(A) = 2/9 = 1/3$ y $P(B) = 4/9$.

Sin embargo, como sabemos que la bolilla es roja, la probabilidad de que sea rayada es $1/2$, ya que, de las rojas la mitad es lisa y la mitad rayada. Observemos, que al ocurrir B, el espacio muestral se *reduce*.

En general, dado un experimento y su espacio muestral asociado, queremos determinar cómo afecta a la probabilidad de A el hecho de saber que ha ocurrido otro evento B.

Definición: Sean A y B eventos tales que $P(B) > 0$, la probabilidad del evento A condicional a la ocurrencia del evento B es

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

Ejemplos: 1) En el ejemplo anterior, $P(B) = 4/9$ y

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/9}{4/9} = \frac{2}{4} = \frac{1}{2}.$$

2) Consideremos una población en la que cada individuo es clasificado según dos criterios: es o no portador de HIV y pertenece o no a cierto grupo de riesgo que denominaremos R. La correspondiente tabla de probabilidades es:

	Portador (A)	No portador (A ^c)	
Pertenece a R (B)	0.003	0.017	0.020
No pertenece a R (B ^c)	0.003	0.977	0.980
	0.006	0.994	1.000

En esta población, la probabilidad de que un individuo sea portador es $P(A) = 0.006$ y la probabilidad de que sea portador y pertenezca al grupo de riesgo R es $P(A \cap B) = 0.003$.

Dado que una persona seleccionada al azar pertenece al grupo de riesgo R, ¿cuál es la probabilidad de que sea portador?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.003}{0.020} = 0.150$$

es decir que 150 de cada 1000 individuos del grupo de riesgo R, son “probablemente” portadores de HIV.

Calculemos ahora la probabilidad de que una persona sea portadora de HIV, dado que no pertenece al grupo de riesgo R.

$$P(A | B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{0.003}{0.980} = 0.00306$$

es decir que sólo 3 de cada 1000 individuos no pertenecientes al grupo de riesgo R, son “posibles” portadores de HIV.

Propiedades de la Probabilidad condicional: Dado un suceso B fijo tal que $P(B) > 0$, $P(\bullet|B)$ es una probabilidad, en el sentido que satisface los axiomas de probabilidad y por lo tanto todas las propiedades que se deducen a partir de ellos. Por ejemplo:

A1. $P(A|B) \geq 0$ para todo suceso A.

A2. $P(S|B) = 1$.

$$\text{Dem: } P(S | B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Ejercicios: 1) Verificar que $P(\bullet|B)$ satisface el axioma **A3a**.

2) Verificar que $P((A_1 \cup A_2) | B) = P(A_1 | B) + P(A_2 | B) - P((A_1 \cap A_2) | B)$

Regla del producto: Dados dos sucesos A y B, tales que $P(B) > 0$,

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

Si además, $P(A) > 0$,

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

Ejemplo: En el ejemplo presentado al comienzo, supongamos ahora que se extraen dos bolillas *sin reposición*. ¿Cuál es la probabilidad de extraer una bolilla roja y una blanca, en ese orden?

Sean C: “la primera bolilla es roja” y D: “la segunda bolilla es blanca”. debemos calcular $P(C \cap D)$. Aplicando la regla del producto

$$P(C \cap D) = P(C) P(D|C) = \frac{4}{9} \frac{5}{8} = \frac{20}{72} = \frac{5}{18}.$$

La regla del producto es especialmente útil cuando el experimento consta de varias etapas ya que se puede generalizar. Así, por ejemplo, si $P(A_1) > 0$ y $P(A_1 \cap A_2) > 0$, se tiene

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2 | A_1) P(A_3 | (A_1 \cap A_2))$$

y se extiende a n sucesos.

Ejemplo: En el mismo ejemplo, ¿cómo podemos obtener la probabilidad de que la segunda bolilla extraída sea blanca (suceso D)?. Sabemos calcular, usando la regla del producto la probabilidad de que la segunda sea blanca y la primera sea roja. Hemos visto que esta probabilidad es $P(C \cap D) = 5/18$. Del mismo modo podemos obtener la probabilidad de que ambas bolillas sean blancas (suceso $(D \cap C^c)$). Esta probabilidad es

$$P(C^c \cap D) = P(C^c) P(D|C^c) = \frac{5}{9} \frac{4}{8} = \frac{20}{72} = \frac{5}{18}.$$

Si ahora observamos que el suceso D puede escribirse como

$$D = (D \cap C) \cup (D \cap C^c)$$

se obtiene

$$P(D) = P(D \cap C) + P(D \cap C^c) = \frac{5}{18} + \frac{5}{18} = \frac{5}{9}. \quad (1)$$

¿Cómo podemos obtener ahora la probabilidad de que la primera bolilla haya sido roja (suceso C) sabiendo que la segunda fue blanca (suceso D)? La probabilidad requerida es

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{5/18}{5/9} = \frac{1}{2}. \quad (2)$$

Los resultados (1) y (2) son ejemplos de aplicación de los dos Teoremas que veremos a continuación: el Teorema de la Probabilidad Total y el Teorema de Bayes, respectivamente.

Definición: Una colección de eventos A_1, A_2, \dots, A_k constituye una *partición* del espacio muestral S si

1. $A_i \cap A_j = \emptyset \quad \forall i \neq j$
2. $P(A_i) > 0 \quad \forall i$
3. $\bigcup_{i=1}^k A_i = S$

Teorema de la probabilidad total: Sea A_1, A_2, \dots, A_k una partición del espacio muestral S y sea B un suceso cualquiera,

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

Dem:

$$B = B \cap S = B \cap \left(\bigcup_{i=1}^k A_i \right) = \bigcup_{i=1}^k (B \cap A_i)$$

Como $(B \cap A_i) \cap (B \cap A_j) = \emptyset \quad \forall i \neq j$, entonces

$$P(B) = P\left(\bigcup_{i=1}^k (B \cap A_i) \right) = \sum_{i=1}^k P(B \cap A_i) = \sum_{i=1}^k P(B | A_i) P(A_i).$$

Teorema de Bayes: Sea A_1, A_2, \dots, A_k una partición del espacio muestral S y sea B un suceso cualquiera tal que $P(B) > 0$,

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

Dem:

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

En el numerador se aplicó la regla del producto y en el denominador el Teorema de la probabilidad total.

El Teorema de Bayes describe cómo es posible “revisar” la probabilidad inicial de un evento o probabilidad a priori ($P(A_i)$) para reflejar la información adicional que nos provee la ocurrencia de un evento relacionado. La probabilidad revisada se denomina probabilidad a posteriori.

Ejemplo: Supongamos que cierta prueba para detectar la presencia de una enfermedad en un individuo, da resultado positivo (detecta la presencia de la enfermedad) en un individuo enfermo con probabilidad 0.99 y en un individuo sano con probabilidad 0.02 (*falso positivo*). Por lo tanto, dicha prueba no detecta la enfermedad en un individuo sano con probabilidad 0.98 y no la detecta en un individuo enfermo con probabilidad 0.01 (*falso negativo*). Es decir que si denotamos A: “la persona padece esa enfermedad” y B: “la prueba es positiva”,

$$P(B | A) = 0.99 \quad P(B | A^c) = 0.02$$

$$P(B^c | A) = 0.01 \quad P(B^c | A^c) = 0.98$$

Se supone, en base a estudios previos, que la incidencia de esa enfermedad en cierta población es 0.001, es decir que la probabilidad a priori de A es 0.001. Se selecciona al azar un individuo de esa población, se le aplica la prueba y el resultado es positivo, ¿cuál es la probabilidad de que en efecto padezca la enfermedad?

Debemos calcular la probabilidad a posteriori de A, $P(A|B)$:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | A^c)P(A^c)} = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.02 \cdot 0.999} = 0.0472$$

Por lo tanto, la probabilidad de que esté enfermo, habiendo sido positivo el resultado de la prueba es aproximadamente 0.05.

Las probabilidades a posteriori dependen fuertemente de las probabilidades a priori. Si se aplica la prueba a individuos de una población en la cual la incidencia de la enfermedad es mucho mayor, también aumentará la probabilidad a posteriori.

Verifique ésto, suponiendo ahora que $P(A) = 0.01$.

Más adelante, desarrollaremos otro ejemplo de aplicación de estos Teoremas.

Independencia

La definición de probabilidad condicional nos permite “revisar” la probabilidad $P(A)$ asignada a un suceso, cuando se sabe que otro suceso B ha ocurrido. Hay casos en los que $P(A | B) \neq P(A)$, mientras que en otros $P(A | B) = P(A)$, es decir que la ocurrencia del suceso B no altera la probabilidad de ocurrencia de A.

Ejemplo: De una urna que contiene 4 bolillas negras y 6 blancas se extraen dos bolillas *sin reposición*, ¿cuál es la probabilidad de que la segunda bolilla sea blanca, sabiendo que la primera es negra?

Denominando A : “la segunda bolilla es blanca” y B : “la primera bolilla es negra”,

$$P(A|B) = \frac{6}{9} = \frac{2}{3}.$$

Por otra parte,

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = \frac{6}{9} \frac{4}{10} + \frac{5}{9} \frac{6}{10} = \frac{54}{90} = \frac{6}{10} = \frac{3}{5}$$

y, por lo tanto, $P(A|B) \neq P(A)$, es decir que la ocurrencia del suceso B modifica la probabilidad del suceso A .

Observemos que la probabilidad de que la segunda bolilla sea blanca coincide con la probabilidad de que la primera lo sea.

Ejercicio: Verificar que, en cambio, si las extracciones se realizan con reposición, $P(A) = P(A|B)$.

Diremos que los eventos A y B son *independientes* si la información acerca de la ocurrencia o no de uno de ellos no afecta la probabilidad de ocurrencia del otro,

Definición: Los eventos A y B son independientes si

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

Si la igualdad no se cumple, decimos que A y B son dependientes.

Proposición: Supongamos $P(B) > 0$, A y B son independientes si y sólo si $P(A|B) = P(A)$.

Dem: (\Rightarrow) Si $P(B) > 0 \Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)}$ está bien definida, pero por ser A y B independientes, $P(A \cap B) = P(A)P(B)$, entonces

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

(\Leftarrow) Aplicando la regla del producto, si $P(B) > 0$, $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$.

Observación: Si $P(B) = 0$, como $A \cap B \subseteq B$, $P(A \cap B) = 0$, y por lo tanto la igualdad $P(A \cap B) = P(A)P(B)$ siempre se satisface.

Ejemplo: De un mazo de 40 cartas españolas, se extrae una carta al azar. Consideremos los siguientes sucesos:

A: "la carta es copa o espada"

B: " la carta no es copa"

C: "la carta es copa u oro"

$$P(A) = \frac{20}{40} = \frac{1}{2} \quad P(B) = \frac{30}{40} = \frac{3}{4} \quad P(C) = \frac{20}{40} = \frac{1}{2}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \neq P(A), \text{ entonces } A \text{ y } B \text{ no son independientes.}$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} = P(A), \text{ entonces } A \text{ y } C \text{ son independientes.}$$

Propiedades: 1) Si los sucesos A y B son excluyentes, es decir si $A \cap B = \emptyset$ y si $P(A) > 0$, $P(B) > 0$, entonces A y B no son independientes.

Dem: En efecto, en este caso, $0 = P(A \cap B) \neq P(A)P(B)$.

2) Si $P(B) = 0$, entonces B es independiente de cualquier suceso A tal que $P(A) > 0$.

Dem: Como $A \cap B \subseteq B$, $P(A \cap B) = 0$ y por lo tanto $P(A \cap B) = P(A)P(B)$, es decir que A y B son independientes.

3) Si $A \subseteq B$, $P(A) > 0$ y $P(B) < 1$, A y B no son independientes.

Dem: Como $A \subseteq B \Rightarrow A \cap B = A \Rightarrow P(A \cap B) = P(A) \neq P(A)P(B)$. Luego, A y B no son independientes.

4) Si A y B son sucesos independientes, A y B^c también lo son.

Dem:

$$P(A) = P(A \cap B) + P(A \cap B^c) \Rightarrow P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) =$$

$$P(A)(1 - P(B)) = P(A)P(B^c).$$

Ejercicio: Demostrar que si A y B son sucesos independientes, A^c y B^c también lo son.

Independencia de más de dos eventos: La definición de independencia de dos eventos puede extenderse a más de dos.

Definición: Los eventos A_1, A_2, \dots, A_n son independientes si para todo $k = 2, \dots, n$ y para todo conjunto de índices $\{i_1, i_2, \dots, i_k\}$ tales que $1 \leq i_1 < i_2 < \dots < i_k \leq n$, se verifica

$$P(A_{i_1} \cap A_{i_2} \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \dots P(A_{i_k})$$

Es decir que es necesario verificar $\binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = 2^n - n - 1$ condiciones.

Observación: Si los sucesos A_1, A_2, \dots, A_n son independientes, entonces son independientes de a pares pero la recíproca no es cierta.

Ejemplos: 1) Sea $S = \{w_1, w_2, w_3, w_4\}$ un espacio de equiprobabilidad y consideremos $n = 3$ y los sucesos

$$A = \{w_1, w_4\}$$

$$B = \{w_2, w_4\}$$

$$C = \{w_3, w_4\}$$

$$P(A) = P(B) = P(C) = \frac{1}{2}.$$

Además,

$$P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

$$P(A \cap C) = \frac{1}{4} = P(A)P(C)$$

$$P(B \cap C) = \frac{1}{4} = P(B)P(C)$$

es decir, que los sucesos son independientes de a pares. Sin embargo,

$$P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C)$$

y, por lo tanto, los sucesos A, B y C no son independientes.

2) Veamos un ejemplo también para el caso $n = 3$, en el cual se satisface la factorización de $P(A \cap B \cap C)$ y no se cumple para alguna de las intersecciones dobles. Sea $S = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8\}$ un espacio de equiprobabilidad y consideremos los sucesos

$$A = \{w_1, w_2, w_3, w_4\}$$

$$B = \{w_1, w_2, w_7, w_8\}$$

$$C = \{w_1, w_5, w_6, w_7\}$$

Como antes, $P(A) = P(B) = P(C) = \frac{1}{2}$. Además,

$$P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

$$P(B \cap C) = \frac{1}{4} = P(B)P(C)$$

$$P(A \cap C) = \frac{1}{8} \neq P(A)P(C)$$

Se observa que no se satisface una de las igualdades, pero sí se satisface

$$P(A \cap B \cap C) = \frac{1}{8} = P(A)P(B)P(C).$$

Finalmente, veremos un ejemplo en el que utilizamos los diferentes conceptos y propiedades estudiadas en esta Sección.

Ejemplo: Muchos sistemas de computación trabajan con enormes bases de datos, como por ejemplo, sistemas de tarjetas de crédito o sistemas de reservas de pasajes aéreos. Debido al volumen de datos involucrado, la velocidad de acceso al sistema depende de las características de las unidades de almacenamiento utilizadas, como así también de las redes de comunicación conectadas a la base de datos. Nos concentraremos en el primer aspecto, es decir en el problema del almacenamiento.

Consideremos unidades de almacenamiento consistentes en discos planos, cada uno de los cuáles está compuesto por un conjunto de anillos concéntricos denominados "pistas". Cada pista está a su vez subdivida en áreas de almacenamiento denominadas "sectores". El acceso al disco se realiza mediante una cabeza lectora/grabadora que se puede mover hacia adelante o hacia atrás a lo largo de un brazo fijo. El disco rota bajo ese brazo y la cabeza lee o modifica un dato cuando el correspondiente sector pasa bajo ella. Consideremos un disco que consiste de 76 pistas, numeradas de 0 a 75, con 8 sectores cada una, numerados de 0 a 7.

Supongamos que, en el momento en que se debe acceder a un dato que se encuentra en el sector 2 de la pista 51, la cabeza se encuentra sobre la pista 22. Entonces, debe moverse en primer lugar hasta la pista 51 (este movimiento se llama búsqueda o *seek*) y luego debe esperar hasta que el sector 2 pase bajo ella (este período de tiempo se denomina retardo rotacional o *rotational delay*).

Si el cabezal se mueve por ejemplo a una velocidad de 3.2 milisegundos (ms) por pista, la búsqueda del ejemplo demandaría $(3.2)(51-22) = (3.2)(29) = 92.8$ ms. Si además suponemos que el disco realiza una rotación completa en 30 ms, el retardo rotacional

puede demorar entre 0 y 30 ms, con un promedio de 15 ms. Por último, supongamos que el acceso concreto al dato demora 1.2 milisegundos.

Este sistema es de naturaleza probabilística o aleatoria. Las demandas de acceso arriban en tiempos aleatorios y se demandan datos aleatorios, en el sentido de que no sabemos con anticipación qué dato se va a requerir.

Analicemos el siguiente ejemplo. Supongamos que las probabilidades de que una demanda de acceso corresponda a cada una de las 76 pistas son iguales y que accesos sucesivos son independientes. Supongamos también que la cabeza lectora/grabadora se encuentra sobre la pista 20, ¿cuál es la probabilidad de que el tiempo total de búsqueda (seek) para las dos siguientes demandas de acceso sea a lo sumo 50 ms?

Sea A el suceso "la búsqueda combinada demora a lo sumo 50 ms" y definamos, para cada $i = 0, 1, \dots, 75$, los sucesos

T_i : "el primero de los dos accesos siguientes corresponderá a un dato que está sobre la pista i "

Entonces

$$P(A) = \sum_{i=0}^{75} P(A \cap T_i) = \sum_{i=0}^{75} P(A|T_i)P(T_i) \quad (3)$$

Como se observa, debemos calcular $P(A|T_i)$, es decir debemos calcular la probabilidad de que la búsqueda combinada demore a lo sumo 50 ms dado que el primer acceso es a la pista i , para $i = 0, 1, \dots, 75$. Por ejemplo, ¿cómo calcularíamos $P(A|T_{26})$?

Si la primera búsqueda nos lleva a la pista 26, demandará $(26-20)(3.2)$ ms = 19.2 ms, por lo tanto la búsqueda total llevará a lo sumo 50 ms si la segunda búsqueda demora a lo sumo 30.8 ms. Como en 30.8 ms se pueden recorrer a lo sumo 9 pistas $(30.8/3.2)$, no podemos ir más allá de la pista $26-9=17$ o de la pista $26+9=35$. En otras palabras $P(A|T_{26})$ será la probabilidad de que el segundo pedido de acceso se refiera a un dato que está entre las pistas 17 y 35 inclusive. Dado que suponemos que todas las pistas son equiprobables,

$$P(A|T_{26}) = \frac{19}{76} = \frac{1}{4}.$$

Del mismo modo, se calculan todas las probabilidades condicionales requeridas en (3) y se obtiene el valor de $P(A)$ pedido.