Índice

Tema 2

Probabilidad condicionada. Independencia

2.1 Introducción

Dependiendo de la información que se tenga en un momento determinado, el espacio muestral de un experimento aleatorio puede variar. Cuanta más información se tenga, más se reducirá el conjunto de posibles resultados y por lo tanto las probabilidades de que ocurran los sucesos pueden cambiar. Esto dará lugar a la llamada probabilidad condicionada.

Estos cambios en las probabilidades pueden hacer que el suceso estudiado sea más probable, menos probable e incluso no cambie, en cuyo caso diremos que existe independencia.

Las probabilidades condicionadas, además, facilitan el cálculo de la probabilidad de algunos sucesos utilizando la información de sucesos que han ocurrido o pudieran ocurrir. Así, el teorema de la probabilidad total y el teorema de Bayes serán especialmente útiles cuando el experimento sigue un cierto proceso por etapas, aunque su aplicación es más diversa.

La probabilidad tiene grandes aplicaciones en una amplia variedad de campos, entre ellos, en la Informática. En particular, el teorema de Bayes se aplica al filtrado de correos electrónicos basura o no deseados (*spam*) tan frecuentes en nuestros días.

En todo este tema supondremos que trabajamos con un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ asociado a un experimento aleatorio, es decir, un espacio muestral, Ω, un álgebra o un sigma álgebra, $\mathcal{A}$, y una probabilidad, P.

2.2 Probabilidad condicionada

En las probabilidades definidas en el tema anterior no se suponen condiciones especiales aparte de las que se definen en el propio experimento aleatorio. Pero, a veces, se dispone de información adicional que puede ser aprovechada para conocer la probabilidad de que ocurra un suceso en situaciones concretas.

Así, si lanzamos un dado, la probabilidad de que salga un cinco es $1/6$. Pero si nos dicen que hemos obtenido un resultado impar, la probabilidad de que salga un cinco es superior, $1/3$.

Otro ejemplo se basa en el hecho de que los *spammers* aprovechan los fines de semana para enviar mayor cantidad de correos electrónicos con publicidad. Supongamos que, en general, cierta persona recibe entre su correo un 20% de correo *spam*, pero en fin de semana este porcentaje asciende al 30%. Esto nos dice que, en general, la probabilidad de recibir *spam* es 0.2, pero la probabilidad de recibir spam un día del fin de semana es 0.3.

En general, dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, queremos estudiar cómo influye la ocurrencia de un suceso B en la probabilidad de que ocurra otro suceso A.

Definición 2.1 *Dado un suceso B con $\mathrm{P}[B] > 0$, se define la probabilidad de un suceso A condicionada a B, como:*

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]} \,. \tag{2.1}$$

La probabilidad condicionada anterior da solución a la siguiente pregunta: sabiendo que B ha ocurrido, ¿cuál es la probabilidad de que ocurra A?

Obsérvese que la definición no tendría sentido si $\mathrm{P}[B] = 0$.

La expresión (2.1) permite hallar la probabilidad condicionada a partir de las probabilidades calculadas considerando todo el espacio muestral inicial Ω.

Desde el punto de vista práctico, cuando se condiciona a un suceso, realmente se produce una reducción del espacio muestral: se sabe no sólo que el resultado obtenido es un elemento del espacio muestral sino que concretamente está en B, por lo que el conjunto de posibles resultados de los que ahora se parte es B.

Para ver que la definición de probabilidad condicionada que se ha proporcionado realmente corresponde a la probabilidad de un suceso A suponiendo que B ha ocurrido, supongamos, por ejemplo, que el espacio muestral es finito y los sucesos elementales son equiprobables. En tal caso:

$$\mathrm{P}[A|B] = \frac{n^o \textit{ de casos favorables}}{n^o \textit{ de casos posibles}} = \frac{card(A \cap B)}{card(B)} = \frac{card(A \cap B)/card(\Omega)}{card(B)/card(\Omega)} = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}$$

Una interpretación similar se puede hacer con la definición frecuentista ya que las frecuencias relativas verifican:

$$f_{A|B} = \frac{f_{A \cap B}}{f_B}.$$

Veamos los ejemplos anteriores escritos formalmente.

Ejemplo 2.1 *Si lanzamos un dado, la probabilidad de que salga un cinco es*

$$\mathrm{P}\left[\{5\}\right] = 1/6,$$

ya que $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Sin embargo, si sabemos que hemos obtenido un resultado impar, la probabilidad de que haya salido un cinco es

$$\mathrm{P}\left[\{5\} \,|\, \{1, 3, 5\}\right] = 1/3.$$

En este ejemplo tendríamos dos sucesos $A = \{5\}$ *y* $B = \{1, 3, 5\}$. *Si sabemos que el resultado obtenido ha sido impar, es decir, ha ocurrido* B*, en realidad el espacio muestral, conjunto de los posibles resultados, se ha reducido a* $\Omega^* = B = \{1, 3, 5\}$.

También podemos calcular la probabilidad anterior utilizando el espacio muestral inicial:

$$\mathrm{P}\left[\{5\} \,|\, \{1, 3, 5\}\right] = \frac{\mathrm{P}\left[\{5\} \cap \{1, 3, 5\}\right]}{\mathrm{P}\left[\{1, 3, 5\}\right]} = \frac{\mathrm{P}\left[\{5\}\right]}{\mathrm{P}\left[\{1, 3, 5\}\right]} = \frac{1/6}{3/6} = \frac{1}{3} = 1/3.$$

Ejemplo 2.2 *Si cierta persona, en general, recibe entre su correo un 20% de correo* spam*, pero en fin de semana este porcentaje asciende al 30%, entonces dados los sucesos* $A =$ *"un correo seleccionado al azar entre los recibidos en un día sea spam" y* $B =$ *"se seleccione un correo al azar recibido en un día del fin de semana", se verifica que:*

$$\mathrm{P}\left[A\right] = 0.2,$$
$$\mathrm{P}\left[A|B\right] = 0.3.$$

Veamos ahora un nuevo ejemplo.

Ejemplo 2.3 *En una ciudad se estudia la cantidad de usuarios de internet según el sexo. Supongamos que se tienen los siguientes datos, en miles de individuos:*

	Hombre	*Mujer*	*Total*
Usa internet	*40*	*35*	*75*
No usa internet	*185*	*240*	*425*
Total	*225*	*275*	*500*

Si seleccionamos un ciudadano al azar y consideramos los sucesos I: "usar internet"; H: "ser hombre", se verifica que

$$\mathrm{P}\left[I\right] = \frac{75}{500} = 0.15 \qquad y \qquad \mathrm{P}\left[H\right] = \frac{225}{500} = 0.45.$$

Además, $\mathrm{P}\left[I \cap H\right] = \frac{40}{500} = 0.08$.

Si mediante un procedimiento aleatorio seleccionamos a un ciudadano varón, ¿cuál es la probabilidad de que use Internet?

Según la intuición, como ya sabemos que es un hombre, nos restringimos al conjunto de 225 hombres (cambiamos de espacio muestral) y, entonces,

$$\mathrm{P}\left[I|H\right] = \frac{40}{225} = 0.1\widehat{7}.$$

Si aplicáramos la definición de probabilidad condicionada,

$$\mathrm{P}[I|H] = \frac{\mathrm{P}[I \cap H]}{\mathrm{P}[H]} = \frac{0.08}{0.45} = 0.1\widehat{7}.$$

obteniendo el mismo resultado.

Teorema 2.1 $\mathrm{P}[\cdot\,|B]$ *es una probabilidad, es decir,*

1. *Para cualquier suceso A, $0 \leq \mathrm{P}[A|B]$.*
2. $\mathrm{P}[\Omega|B] = 1$.
3. *En el caso en el que $\mathcal{A}$ sea un álgebra, dados A_1, A_2 dos sucesos mutuamente excluyentes $A_1 \cap A_2 = \emptyset$, se verifica que*

$$\mathrm{P}[A_1 \cup A_2|B] = \mathrm{P}[A_1|B] + \mathrm{P}[A_2|B].$$

En el caso en el que $\mathcal{A}$ sea un sigma-álgebra, dados $A_1, A_2, \ldots,$ sucesos mutuamente excluyentes, se verifica que

$$\mathrm{P}\left[\bigcup_{i=1}^{\infty} A_i|B\right] = \sum_{i=1}^{\infty} \mathrm{P}[A_i|B].$$

Como consecuencia, $\mathrm{P}[\cdot\,|B]$ verifica todas las propiedades de una probabilidad. En particular, $\mathrm{P}\left[\bar{A}|B\right] = 1 - \mathrm{P}[A|B]$.

2.3 Regla de la multiplicación

De la definición de probabilidad condicionada dada en (2.1) se deduce una forma de calcular la probabilidad de la intersección de dos sucesos:

$$\mathrm{P}[A \cap B] = \mathrm{P}[B]\,\mathrm{P}[A|B],$$

si $\mathrm{P}[B] > 0$.

Ejemplo 2.4 *Una caja contiene 10 fusibles, de los cuales 3 vienen defectuosos y los 7 restantes se encuentran en buen estado. Extraemos de forma sucesiva dos fusibles de la caja. Se desea calcular la probabilidad de que el primer fusible extraído sea defectuoso, y el segundo no defectuoso.*

Consideramos los sucesos

D_i : el i-ésimo fusible extraído sea defectuoso, $i = 1, 2$.

N_i : el i-ésimo fusible extraído sea no defectuoso, $i = 1, 2$.

Entonces,

$$\mathrm{P}[D_1 \cap N_2] = \mathrm{P}[D_1]\,\mathrm{P}[N_2|D_1] = \frac{3}{10} \times \frac{7}{9} = \frac{7}{30} = 0.2\widehat{3}.$$

El resultado anterior se puede generalizar a tres o más sucesos. Así, si $\mathrm{P}[A_1 \cap A_2] > 0$.

$$\mathrm{P}[A_1 \cap A_2 \cap A_3] = \mathrm{P}[A_1]\,\mathrm{P}[A_2|A_1]\,\mathrm{P}[A_3|A_1 \cap A_2].$$

Obsérvese que si $\mathrm{P}[A_1 \cap A_2] > 0$, entonces $\mathrm{P}[A_1] \geq \mathrm{P}[A_1 \cap A_2] > 0$, por lo que tiene sentido calcular la probabilidad condicionada $\mathrm{P}[A_2|A_1]$.

Teorema 2.2 *Dados* $A_1, A_2, \ldots, A_n$ *sucesos, si* $\mathrm{P}\left[A_1 \cap A_2 \cap \ldots \cap A_{n-1}\right] > 0,$ *entonces*

$$\mathrm{P}\left[A_1 \cap A_2 \cap \ldots \cap A_n\right] = \mathrm{P}\left[A_1\right] \mathrm{P}\left[A_2 | A_1\right] \cdot \ldots \cdot \mathrm{P}\left[A_n | A_1 \cap A_2 \cap \ldots \cap A_{n-1}\right].$$

Ejemplo 2.5 *Suponemos que en el ejemplo anterior, en el que teníamos 3 fusibles defectuosos y 7 en buen estado, extraemos sucesivamente cuatro fusibles y nos planteamos calcular la probabilidad de que el primero sea defectuoso y el resto no lo sea.*

La probabilidad que queremos calcular es $\mathrm{P}\left[D_1 \cap N_2 \cap N_3 \cap N_4\right]$. *Utilizando la regla de la multiplicación, podemos obtenerla de la siguiente manera:*

$$\begin{aligned} \mathrm{P}\left[D_1 \cap N_2 \cap N_3 \cap N_4\right] &= \mathrm{P}\left[D_1\right] \mathrm{P}\left[N_2 | D_1\right] \mathrm{P}\left[N_3 | D_1 \cap N_2\right] \mathrm{P}\left[N_4 | D_1 \cap N_2 \cap N_3\right] = \\ &= \frac{3}{10} \times \frac{7}{9} \times \frac{6}{8} \times \frac{5}{7} = 0.125. \end{aligned}$$

2.4 Independencia de sucesos

De la definición de probabilidad condicionada dada en (2.1) y los ejemplos anteriores vemos que en general, la $\mathrm{P}\left[A|B\right]$ no tiene porqué ser igual a la $\mathrm{P}\left[A\right]$. De hecho se tienen las siguientes posibilidades:

- $\mathrm{P}\left[A|B\right] > \mathrm{P}\left[A\right]$, se dice que B favorece la aparición de A.
- $\mathrm{P}\left[A|B\right] < \mathrm{P}\left[A\right]$, se dice que B no favorece la aparición de A.
- $\mathrm{P}\left[A|B\right] = \mathrm{P}\left[A\right]$, caso particular que estudiamos a continuación con más detalle.

Definición 2.2 *Dados dos sucesos* A *y* B*, tal que* $\mathrm{P}\left[B\right] > 0$*, se dice que* A *es independiente de* B *si*

$$\mathrm{P}\left[A|B\right] = \mathrm{P}\left[A\right].$$

Es decir, si se conoce que B ha ocurrido, la probabilidad de A es la misma que se tenía inicialmente.

Ejemplo 2.6 *En el ejemplo 2.3, teníamos la tabla*

	Hombre	*Mujer*	*Total*
Usa internet	*40*	*35*	*75*
No usa internet	*185*	*240*	*425*
Total	*225*	*275*	*500*

de donde se deducía que

$$\mathrm{P}\left[I\right] = 0.15; \qquad \mathrm{P}\left[I|H\right] = 0.1\widehat{7}$$

En este caso, el suceso I *no es independiente de* H. *Es más,* $\mathrm{P}\left[I\right] < \mathrm{P}\left[I|H\right]$, *por lo que la probabilidad de usar internet se incrementa cuando nos restringimos a los varones.*

Por otro lado, si estudiamos la probabilidad de que una mujer escogida al azar use internet, tendremos que:

$$\mathrm{P}\left[I|\overline{H}\right] = \frac{35}{275} = 0.1\widehat{27} < \mathrm{P}\left[I\right]$$

lo que nos dice que la probabilidad de usar internet disminuye al restringirnos a las mujeres.

En cambio, si la tabla que teníamos fuera

	Hombre	*Mujer*	*Total*
Usa internet	*90*	*110*	*200*
No usa internet	*135*	*165*	*300*
Total	*225*	*275*	*500*

Entonces,

$$\mathrm{P}\left[I\right] = \frac{200}{500} = 0.4; \qquad \mathrm{P}\left[I|H\right] = \frac{90}{225} = 0.4.$$

En este caso, la probabilidad de usar internet no varía cuando nos restringimos a los ciudadanos varones, por lo que I es independiente de H.

Ejemplo 2.7 *En el ejemplo 2.4, en el que teníamos una caja con 3 fusibles defectuosos y 7 en buen estado, si las extracciones se realizan con reemplazamiento, es decir, tras extraer un fusible lo devolvemos a la caja y, tras revolver los fusibles volvemos a sacar otro, el resultado de la segunda extracción será independiente del resultado de la primera extracción.*

A partir de la probabilidad de una intersección de sucesos se puede establecer una condición equivalente a la definición de independencia ya que si A es independiente de B:

$$\mathrm{P}\left[A \cap B\right] = \mathrm{P}\left[B\right] \mathrm{P}\left[A|B\right] = \mathrm{P}\left[B\right] \mathrm{P}\left[A\right] = \mathrm{P}\left[A\right] \mathrm{P}\left[B\right].$$

Obsérvese que si A es independiente de B y, además, $\mathrm{P}\left[A\right] > 0$, también se verificará que B es independiente de A, ya que

$$\mathrm{P}\left[B \cap A\right] = \mathrm{P}\left[A \cap B\right] = \mathrm{P}\left[A\right] \mathrm{P}\left[B\right] = \mathrm{P}\left[B\right] \mathrm{P}\left[A\right].$$

Por lo que la definición de independencia se puede unificar incluso aunque alguno de los dos sucesos tenga probabilidad nula de ocurrir.

Definición 2.3 *Dados dos sucesos A y B, se dice que A y B son independientes si*

$$\mathrm{P}\left[A \cap B\right] = \mathrm{P}\left[A\right] \mathrm{P}\left[B\right].$$

Este resultado facilita el cálculo de la probabilidad de la intersección de dos sucesos cuando **ambos sucesos** son **independientes**.

Ejemplo 2.8 *Comprobemos de nuevo que los sucesos I y H son independientes para la tabla alternativa del ejemplo anterior (Ejemplo 2.6), ahora utilizando la definición 2.3.*

Por un lado, $\mathrm{P}\left[I \cap H\right] = \frac{90}{500} = 0.18$.

Por otro lado,

$$\mathrm{P}\left[I\right] = 0.4; \qquad \mathrm{P}\left[H\right] = \frac{225}{500} = 0.45 \Longrightarrow \mathrm{P}\left[I\right] \cdot \mathrm{P}\left[H\right] = 0.4 \cdot 0.45 = 0.18.$$

Luego I y H son independientes.

De la independencia de dos sucesos se puede deducir la independencia con sucesos complementarios, lo que se recoge en el siguiente teorema.

Teorema 2.3 *Sean A y B dos sucesos independientes. Entonces también son independientes:*

1. A *y* $\bar{B}$.
2. $\bar{A}$ *y* B.
3. $\bar{A}$ *y* $\bar{B}$.

También podemos definir independencia entre sucesos cuando disponemos de más de dos sucesos.

Definición 2.4 *Sean $A_1, \ldots, A_n$ sucesos. Se dice que son independientes dos a dos si y sólo si*

$$\mathrm{P}\left[A_i \cap A_j\right] = \mathrm{P}\left[A_i\right] \mathrm{P}\left[A_j\right] \ , \qquad \forall i \neq j \ .$$

Definición 2.5 *Sean $A_1, \ldots, A_n$ sucesos. Se dice que son independientes si y sólo si*

$$\mathrm{P}\left[A_i \cap A_j\right] = \mathrm{P}\left[A_i\right] \mathrm{P}\left[A_j\right] \ , \qquad \forall i \neq j \ ,$$

$$\mathrm{P}\left[A_i \cap A_j \cap A_k\right] = \mathrm{P}\left[A_i\right] \mathrm{P}\left[A_j\right] \mathrm{P}\left[A_k\right] \ , \qquad \forall i \neq j \neq k \ ,$$

$$\vdots$$

$$\mathrm{P}\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \mathrm{P}\left[A_i\right] \ .$$

Obsérvese que si un conjunto de sucesos son independientes, también lo son dos a dos, aunque el recíproco no es cierto en general.

2.5 Teorema de la probabilidad total y teorema de Bayes

Recordemos que $\{A_1, \ldots, A_n\}$ constituye un sistema completo de sucesos si son:

- mutuamente excluyentes $(A_i \cap A_j = \emptyset, \forall i \neq j)$,
- exhaustivos $(\bigcup_{i=1}^{n} A_i = \Omega)$.

Como consecuencia de las dos condiciones anteriores se tiene que

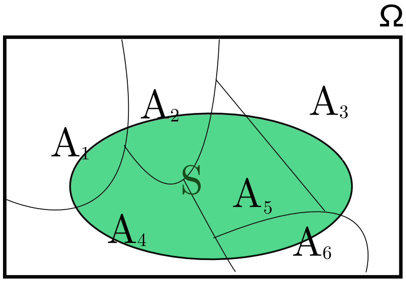
$$\mathrm{P}\left[A_i \cap A_j\right] = 0, \forall i \neq j \qquad y \qquad \sum_{i=1}^{n} \mathrm{P}\left[A_i\right] = 1.$$

El teorema de la probabilidad total permite calcular probabilidades de sucesos aprovechando las probabilidades condicionadas a los distintos elementos de un sistema completo de sucesos. Esto es especialmente útil cuando un experimento consta de dos etapas: para conocer la probabilidad de que ocurra algo en la segunda etapa se condiciona a cada uno de los sucesos que podría haber ocurrido en la primera etapa.

Teorema 2.4 *(**Teorema de la probabilidad total (TPT)**) Dado un sistema completo de sucesos, $\{A_1, \ldots, A_n\}$, con $\mathrm{P}[A_i] > 0$, $i = 1, \ldots, n$, y S otro suceso cualquiera, se tiene*

$$\mathrm{P}[S] = \sum_{i=1}^{n} \mathrm{P}[A_i]\,\mathrm{P}[S|A_i]\,.$$

El siguiente dibujo puede ser útil para comprender la demostración del teorema:



Demostración

El suceso S se puede escribir como $S = S \cap \Omega$.

Al ser los A_i exhaustivos, haciendo uso de la propiedad distributiva de la unión e intersección:

$$S = S \cap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} (S \cap A_i)\,.$$

Y por ser mutuamente excluyentes y utilizando la definición de probabilidad condicionada:

$$\mathrm{P}[S] = \mathrm{P}\left[\bigcup_{i=1}^{n} (S \cap A_i) \right] = \sum_{i=1}^{n} \mathrm{P}[S \cap A_i] = \sum_{i=1}^{n} \mathrm{P}[A_i]\,\mathrm{P}[S|A_i]\,.$$

□

Obsérvese que en particular, si A es un suceso tal que $0 < \mathrm{P}[A] < 1$, entonces:

$$\mathrm{P}[S] = \mathrm{P}[A]\,\mathrm{P}[S|A] + \mathrm{P}\left[\overline{A}\right]\mathrm{P}\left[S|\overline{A}\right],$$

ya que A y $\overline{A}$ constituyen un sistema completo de sucesos.

Ejemplo 2.9 *Suponemos que tenemos una urna con 3 bolas blancas y 5 bolas negras, de la que extraemos una bola al azar y, sin devolverla a la urna, realizamos una segunda extracción. Nos preguntamos cuál es la probabilidad de que la segunda bola sea blanca.*

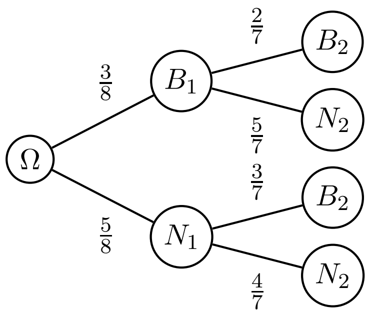
Para resolver el problema, definimos los siguientes sucesos:

B_i : *"se extrae una bola blanca en la i-ésima extracción", $i = 1, 2$.*

N_i : *"se extrae una bola negra en la i-ésima extracción", $i = 1, 2$.*

La probabilidad que nos planteamos es $\mathrm{P}[B_2]$. El problema es que, si no sabemos cuál ha sido el resultado de la primera extracción (primera etapa), ¿cómo podemos calcular la probabilidad de algo que ocurre en la segunda extracción (segunda etapa)?

Gracias al TPT podemos resolver el problema. Representemos en un esquema los posibles resultados de la primera etapa y, a partir de ellos, lo que puede ocurrir en la segunda etapa.



Los números que aparecen en las ramas son las probabilidades de que ocurra cada uno de los sucesos, a partir del resultado anterior. Es decir:

$$\begin{aligned} \mathrm{P}[B_1] &= 3/8, \quad \mathrm{P}[N_1] = 5/8, \\ \mathrm{P}[B_2|B_1] = 2/7, \quad \mathrm{P}[N_2|B_1] &= 5/7, \quad \mathrm{P}[B_2|N_1] = 3/7, \quad \mathrm{P}[N_2|N_1] = 4/7. \end{aligned}$$

$\{B_1, N_1\}$ *constituye un sistema completo de sucesos ya que son* B_1 *y* N_1 *mutuamente excluyentes y exhaustivos (de hecho, son complementarios).*

Utilizando el TPT, podemos calcular $\mathrm{P}[B_2]$ *de la siguiente forma:*

$$\mathrm{P}[B_2] = \mathrm{P}[B_1]\,\mathrm{P}[B_2|B_1] + \mathrm{P}[N_1]\,\mathrm{P}[B_2|N_1] = \frac{3}{8} \times \frac{2}{7} + \frac{5}{8} \times \frac{3}{7} = 0.375$$

Obsérvese que para el cálculo de $\mathrm{P}[B_2]$ *hemos tomado todos los caminos por los que se llega a* B_2 *del esquema, multiplicando las probabilidades que aparecen en cada camino y sumando cada uno de los caminos.*

Ejemplo 2.10 *Parte de la actividad de cierta empresa es construir placas base para ordenadores. Con el objetivo de mantener ciertos niveles de calidad, se inspeccionan cada una de las placas base. De las placas fabricadas, el 60% se destinan a una marca de primera calidad, el 25% se destinan a una marca de segunda calidad y el 15% restante a tercera calidad. Además se conoce la información sobre las placas que han sufrido reclamaciones por funcionamiento inadecuado en un periodo inferior a tres años, resultando de un 10% para las de primera calidad, un 15% para las placas de segunda calidad y un 20% para los de tercera.*

Nos planteamos: entre todas las placas fabricadas, ¿cuál es la probabilidad de que una placa escogida al azar sufra una reclamación antes de tres años?

En este ejemplo a cada placa base se le da un único destino: marca de primera, segunda y tercera calidad. Luego si escogemos al azar una placa base de la empresa y definimos los sucesos:

A_1*="destinarse a una marca de primera calidad",*

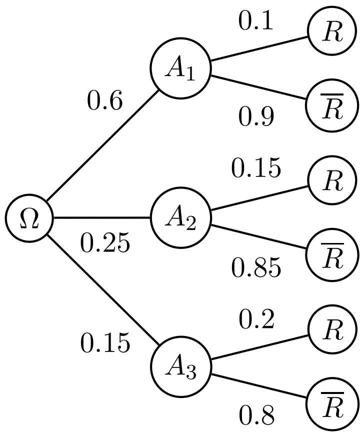
A_2*="destinarse a una marca de segunda calidad",*

A_3*="destinarse a una marca de tercera calidad",*

los tres sucesos son mutuamente excluyentes (no se le puede dar dos usos simultáneos) y exhaustivos (abarcan todos los posibles usos que propone la empresa), por lo que $\{A_1, A_2, A_3\}$ *constituye un sistema completo de sucesos. Entonces, si definimos el suceso* R : *"sufrir una reclamación en un periodo inferior a tres años", entonces, la información de la que disponemos nos permite conocer las siguientes* *probabilidades:*

$$\begin{aligned} \mathrm{P}[A_1] &= 0.6; \qquad \mathrm{P}[A_2] = 0.25; \qquad \mathrm{P}[A_3] = 0.15; \\ \mathrm{P}[R|A_1] &= 0.1; \qquad \mathrm{P}[R|A_2] = 0.15; \qquad \mathrm{P}[R|A_3] = 0.2; \end{aligned}$$

La información que se nos proporciona se puede resumir con el siguiente esquema tipo árbol:



Y podemos resolver la pregunta planteada mediante el TPT:

$$\mathrm{P}[R] = \sum_{i=1}^{3} \mathrm{P}[A_i]\,\mathrm{P}[R|A_i] = 0.6 \times 0.1 + 0.25 \times 0.15 + 0.15 \times 0.2 = 0.1275.$$

Por lo tanto, el 12.75% de las placas fabricadas sufrirán una reclamación antes de 3 años.

Veamos ahora otro tipo de preguntas que podemos plantearnos con ayuda de los ejemplos anteriores:

Si en el ejemplo 2.9 (urna con bolas) vemos que el resultado de la segunda extracción ha sido una bola blanca, ¿cuál es la probabilidad de que en la primera extracción la bola fuera del mismo color?

Si en el ejemplo 2.10 (placas base) se toma una reclamación al azar, ¿cuál es la probabilidad de que corresponda a una placa de primera calidad?

El teorema de Bayes podrá dar solución a estas preguntas. Este teorema es muy útil cuando se realizan experimentos que consisten en dos etapas y se pregunta sobre la probabilidad de algún suceso relativo a la primera etapa cuando se tiene información sobre lo ocurrido en la segunda etapa.

Teorema 2.5 ***(Teorema de Bayes)*** *Dado un sistema completo de sucesos, $\{A_1, \ldots, A_n\}$, con $\mathrm{P}[A_i] > 0$, $i = 1, \ldots, n$, y S otro suceso cualquiera con $\mathrm{P}[S] > 0$, se tiene*

$$\mathrm{P}[A_j|S] = \frac{\mathrm{P}[A_j]\,\mathrm{P}[S|A_j]}{\sum_{i=1}^{n} \mathrm{P}[A_i]\,\mathrm{P}[S|A_i]}$$

Demostración

Por la definición de probabilidad condicionada y utilizando la propiedad conmutativa de la intersección:

$$\mathrm{P}[A_j|S] = \frac{\mathrm{P}[A_j \cap S]}{\mathrm{P}[S]} = \frac{\mathrm{P}[S \cap A_j]}{\mathrm{P}[S]}.$$

Utilizando la regla de la multiplicación en el numerador y desarrollando el denominador mediante el TPT se tiene que:

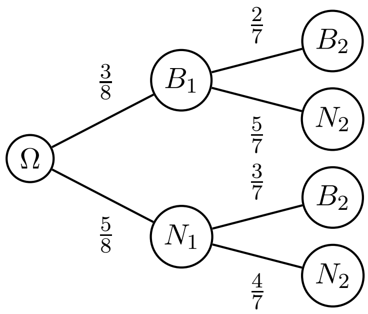
$$\mathrm{P}[A_j|S] = \frac{\mathrm{P}[A_j]\,\mathrm{P}[S|A_j]}{\sum_{i=1}^{n} \mathrm{P}[A_i]\,\mathrm{P}[S|A_i]}$$

□

Ejemplo 2.11 *Si en el ejemplo 2.9, en el que teníamos una urna con 3 bolas blancas y 5 negras tras dos extracciones sucesivas vemos que el resultado de la segunda extracción ha sido una bola blanca, nos planteamos cuál es la probabilidad de que en la primera extracción la bola fuera del mismo color, tendremos que calcular:*

$$\mathrm{P}[B_1|B_2] = \frac{\mathrm{P}[B_1]\,\mathrm{P}[B_2|B_1]}{\mathrm{P}[B_1]\,\mathrm{P}[B_2|B_1] + \mathrm{P}[N_1]\,\mathrm{P}[B_2|N_1]}.$$

Obsérvese que, por el TPT, el denominador coincide con $\mathrm{P}[B_2]$ *que ya ha sido calculado anteriormente, por lo que, según el esquema:*

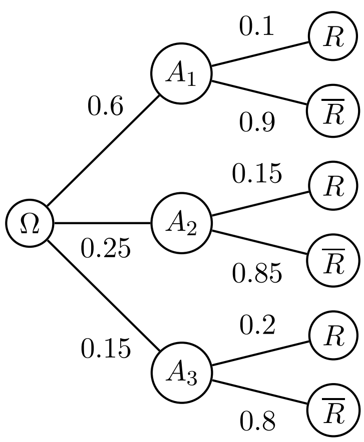


la probabilidad que buscamos es:

$$\mathrm{P}[B_1|B_2] = \frac{\frac{3}{8} \times \frac{2}{7}}{0.375} = 0.2857.$$

Ejemplo 2.12 *Si en el ejemplo 2.10 (placas base) se toma una reclamación al azar, ¿cuál es la probabilidad de que corresponda a una placa de primera calidad?; ¿cuál es el tipo de placa que tiene mayor probabilidad de ser escogido al tomarse una reclamación?*

El esquema que teníamos para este ejemplo era:



Nosotros conocemos los porcentajes fabricados de cada uno de los tipos de placa (primera etapa: fabricación) y dentro de ellos el porcentaje reclamado (segunda etapa: reclamación). Pero en el problema tenemos una reclamación y nos piden la probabilidad de que se haya utilizado como placa de primera calidad, es decir, sabemos el resultado de la segunda etapa y vamos a calcular la probabilidad de lo que haya ocurrido en la primera. Este es un ejemplo característico del teorema de Bayes:

$$\begin{aligned}
\mathrm{P}[A_1|R] &= \frac{\mathrm{P}[A_1]\,\mathrm{P}[R|A_1]}{\sum_{i=1}^{3}\mathrm{P}[A_i]\,\mathrm{P}[R|A_i]} = \frac{\mathrm{P}[A_1]\,\mathrm{P}[R|A_1]}{\mathrm{P}[R]} = \frac{0.6\times 0.1}{0.1275} = 0.4706. \\
\mathrm{P}[A_2|R] &= \frac{\mathrm{P}[A_2]\,\mathrm{P}[R|A_2]}{\sum_{i=1}^{3}\mathrm{P}[A_i]\,\mathrm{P}[R|A_i]} = \frac{\mathrm{P}[A_2]\,\mathrm{P}[R|A_2]}{\mathrm{P}[R]} = \frac{0.25\times 0.15}{0.1275} = 0.2144. \\
\mathrm{P}[A_3|R] &= \frac{\mathrm{P}[A_3]\,\mathrm{P}[R|A_3]}{\sum_{i=1}^{3}\mathrm{P}[A_i]\,\mathrm{P}[R|A_i]} = \frac{\mathrm{P}[A_3]\,\mathrm{P}[R|A_3]}{\mathrm{P}[R]} = \frac{0.15\times 0.2}{0.1275} = 0.2353.
\end{aligned}$$

Estos resultados nos dicen que el 47.06% de las reclamaciones que se hicieron fueron de placas de primera calidad, el 21.44% de segunda calidad y el 23.53% de tercera calidad. Aunque esto aparentemente pueda parecer una contradicción (mayor porcentaje de reclamaciones en las placas de primera calidad), téngase en cuenta que el 60% de las placas base que se fabrican son de primera calidad, por lo que, aunque el porcentaje de reclamaciones para este tipo de placas sea inferior al resto, en el global de las reclamaciones hay un mayor porcentaje que vienen de placas de primera calidad.

2.6 Aplicaciones de la probabilidad a la Informática

El teorema de Bayes ha tenido grandes aplicaciones a lo largo de la historia en distintos campos. Actualmente tiene una gran aplicación en Física (mecánica cuántica), Ingeniería de Telecomunicaciones (señales), control de tráfico en redes, en visión artificial, sistemas expertos, reconocimiento de lenguaje natural, etc. También en los campos de la Informática y la Tecnología se está utilizando cada vez más: en los buscadores, en las cámaras digitales, en los traductores automáticos, los diccionarios de los teléfonos móvil, los filtros para el *spam* en el correo electrónico.

Uno de los aspectos que hemos nombrado es el correo *spam* o no deseado. El correo electrónico se ve últimamente saturado por este tipo de mensajes. Muchos servidores utilizan sistemas para filtrar este correo no deseado. Uno de ellos se basa en el teorema de Bayes (véase, Paul Graham 2002,

`http://www.paulgraham.com/spam.html`).

Este sistema se basa en ciertas puntuaciones que se le asignan a los correos: se tiene una base de datos de las palabras típicas que aparecen en los correos no deseados y se comprueba si aparecen en los correos que llegan al servidor. Muchas de estas palabras pueden aparecer en mensajes que no deseamos eliminar, pero gracias al teorema de Bayes se puede tener en cuenta el contexto de las palabras que aparecen. Cuando se encuentran ciertas palabras el filtro bayesiano calcula la probabilidad de que un correo electrónico sea *spam*. Por ejemplo si en el texto aparece la palabra *viagra*, la probabilidad de que aparezca también la palabra *rich* es muy alta en los spam. Por ello, en el caso de que aparezcan simultáneamente ambas palabras se le asignará una alta puntuación al correo y posiblemente junto con otras puntuaciones, sea descartado como *spam*.

Ejemplo 2.13 *Por ejemplo este correo fue filtrado por el servidor de la Universidad de Sevilla, cuyo asunto era:*

"CONGRATULATIONS!!!YOU HAVE WON"

------------ Comienzo de los resultados de SpamAssassin -------------- Este correo probablemente es spam. El mensaje original ha sido marcado para que pueda reconocer o bloquear en el futuro correo no solicitado, usando las capacidades de filtrado que incorpora su lector de correo.

Detalles del análisis: (15.1 puntos, 8.0 requeridos)

1.0 FROM_ENDS_IN_NUMS From: finaliza en números

1.5 YOU_WON BODY: Who really wins?

1.8 UNCLAIMED_MONEY BODY: People just leave money laying around

5.4 BAYES_99 BODY: Bayesian spam probability is 99 to 100%

[score: 0.9961]

1.2 MAILTO_TO_SPAM_ADDR URI: Includes a link to a likely spammer email

1.5 RCVD_IN_BL_SPAMCOP_NET RBL: Recibido desde un relay en bl.spamcop.net

[Blocked - see <http://www.spamcop.net/bl.shtml?199.181.134.43 >]

2.0 NIGERIAN_BODY1 Message body looks like a Nigerian spam message 1+

0.7 PLING_PLING El Subject tiene muchos signos de exclamación

---------------- Fin de los resultados de SpamAssassin ----------------

A la izquierda de cada una de los elementos examinados que ha detectado indica la puntuación asignada y si la suma de todas ellas (15.1 puntos en el ejemplo) supera un cierto número de puntos (8 puntos en el ejemplo) el correo es calificado como spam. Se puede observar que por probabilidad bayesiana se ha concluido que tiene un 99.61% de probabilidad de ser correo spam, asignando por ello 5.4 puntos.

En gran cantidad de ocasiones los "spammers" *tratan de modificar el texto para que no sea detectado por estos filtros. Por ejemplo es típico el cambio de la palabra* viagra *por \ /llGRA u otras variantes. Este sistema basado en el teorema de Bayes tiene la gran ventaja de adaptarse y aprender sobre las modificaciones introducidas en los spams. No sólo se filtran según las palabras sino que además se usa estadística predictiva.*