1 Probabilidad condicional

Definición 1.1 Asumiendo $P(B) > 0$, definimos la probabilidad condicional de A, dado que B a ocurrido como sigue

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

Para cualquier evento B fijo tal que $P(B) > 0, \mathbb{P}(\cdot|B)$ es una función de probabilidad (es decir, que cumple los tres axiomas de probabilidad). En particular $\mathbb{P}(A|B) \geq 0, \mathbb{P}(\Omega|B) = 1$ y si $A_1, A_2, \ldots$ son disjuntos, entonces $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$. Pero esto no es cierto en general que $\mathbb{P}(A|B \cup C) = \mathbb{P}(A|B) + P(A|C)$.

Las reglas de la probabilidad se aplican a los eventos de la izquierda de la barra $|$. En general, no se cumple que $P(A|B) = P(B|A)$.

Hay mucha confusión con esto todo el tiempo. Por ejemplo, la probabilidad de que tengas puntos (manchas) dado que tienes sarampión es 1, pero la probabilidad de que tengas sarampión, dado que tienes puntos (manchas) no es 1. En este caso, la diferencia entre $\mathbb{P}(A|B)$ y $\mathbb{P}(B|A)$ es obvia, pero hay casos en los que es menos obvio. Este error sucede con bastante frecuencia en los casos legales que a veces se llaman *falacias del fiscal*.

Si $\mathbb{P}(A) > 0$ y $\mathbb{P}(B) > 0$, tenemos

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

Intuitivamente, $\mathbb{P}(A|B)$ es la probabilidad de que A ocurra suponiendo que ocurrió el evento B. De acuerdo con esa intuición, la probabilidad condicional tiene las siguientes propiedades

- Si $B \subset A$ entonces $\mathbb{P}(A|B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.
- Si $A \cap B = \emptyset$ entonces $\mathbb{P}(A|B) = 0/\mathbb{P}(B) = 0$.
- $A \subset B$ entonces $\mathbb{P}(A|B) = \mathbb{P}(A)/\mathbb{P}(B)$.
- La probabilidad condicional, puede ser vista como una función de probabilidad

$$\mathbb{P}_A(E) = \mathbb{P}(E|A)$$

  y todas las propiedades y intuiciones que se aplican a la función probabilidad se aplican a $\mathbb{P}_A$ también.
- Asumiendo que el evento A ocurrió, $\mathbb{P}_A$ generalmente tiene mejores habilidades de pronóstico que $\mathbb{P}$.

Como se mencionó anteriormente, las probabilidades condicionales son usualmente intuitivas. El siguiente ejemplo de (Feller, 1968), sin embargo, muestra una situación contra-intuitiva que implica probabilidades condicionales. Esto demuestra que la intuici?n no debe ser un sustituto de la computación rigurosa.

Ejemplo 1.1 Consideramos dos familias con dos hijos donde la probabilidad de género de cada niño es simétrica $(1/2)$. Seleccionamos una familia al azar y consideramos el espacio muestral que describe el género de los niños $\Omega = \{MM, MF, FM, FF\}$. Asumimos un modelo clásico, implicando que las probabilidades de todos los eventos elementales son $1/4$.

Definimos el evento en que ambos hijos en la familia son niños como $A = \{MM\}$, el evento que una familia tiene un niño es $B = \{MF, FM, MM\}$ y el evento que el primer hijo es un niño como $C = \{MF, MM\}$.

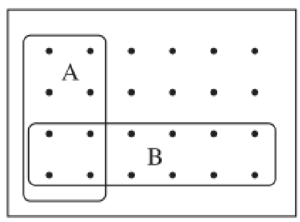
Dado que el primer hijo es un niño, la probabilidad de que ambos hijos sean niños es

$$\mathbb{P}(A|C) = \mathbb{P}(A \cap C)/\mathbb{P}(C) = \mathbb{P}(A)/\mathbb{P}(C) = (1/4)/(1/2) = 1/2.$$

Esto coincide con nuestra intuición. Dado que la familia tiene un niño, la probabilidad de que ambos hijos sean niños es contraintuitiva

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A)/\mathbb{P}(B) = (1/4)/(3/4) = 1/3.$$

Ejemplo 1.2 Observamos en la siguiente figura, que $\mathbb{P}(A \cap B) = 1/6$, que $\mathbb{P}(B) = 1/2$ y que $\mathbb{P}(A|B) = (1/6)/(1/2) = 1/3$, como se esperaba. Sabemos que B ha ocurrido y que el evento A ocurrirá si uno de los cuatro resultados en $A \cap B$ se elige entre los 12 igualmente probable.



Podemos generalizar la ecuación de la probabilidad condicional a múltiples eventos. Sea $A_i, i = 1, 2, \ldots, k$ para k eventos para el cual $\mathbb{P}(A_1 \cap A_2 \cap \cdots A_k) > 0$. Entonces

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots A_k) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

Ejemplo 1.3 En una clase de nivel de posgrado de primer año de 60 estudiantes, diez estudiantes son estudiantes de pregrado.

Calculemos la probabilidad de que tres estudiantes elegidos al azar sean estudiantes de pregrado. Para ello sea A_1 el evento en el cual el primer estudiante elegido es un estudiante de pregrado, A_2 el evento en que el segundo estudiante sea de pregrado y así sucesivamente. Luego de la ecuación anterior tenemos

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)$$

reemplazando obtenemos

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{10}{60} \times \frac{9}{59} \times \frac{8}{58} = 0.003507.$$

Ejemplo 1.4 Un test médico para una enfermedad D tiene salidas + y -

Salidas	D	D^c
+	.009	0.099
-	.001	0.891

Desde la definición de probabilidad condicional,

$$\mathbb{P}(+|D) = \frac{\mathbb{P}(+\cap D)}{\mathbb{P}(D)} = \frac{.009}{.009+.001} = .9$$

y

$$\mathbb{P}(-|D^c) = \frac{\mathbb{P}(-\cap D^c)}{\mathbb{P}(D^c)} = \frac{.891}{.891+.099} \approx .9$$

Al parecer, la prueba es bastante exacta. Personas enfermas producen un positivo del 90 por ciento de la veces y las personas sanas producen una negativa sobre 90 por ciento de las veces . Supongamos que una persona se hace una prueba y obtiene un resultado positivo. ¿ Cuál es la probabilidad de que esa persona tiene la enfermedad? Muchas personas responden que $0,90$, sin embargo la respuesta correcta es

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(+\cap D)}{\mathbb{P}(+)} = \frac{.009}{.009+.099} \approx .08$$

Ejemplo 1.5 Una pequeña empresa de software emplea 4 programadores. El porcentaje de código escrito por cada programador y el porcentaje de errores en su código se muestran en la siguiente tabla. El código es seleccionado al azar.

Programador	% de código escrito	% de errores en código
1	15	5
2	20	3
3	25	2
4	40	1

1. Antes de que el código sea examinado, ¿ cuál es la probabilidad de que fue escrito por el

```
Programador 1? .15
Programador 2? .20
Programador 3? .25
Programador 4? .40
```

Aplicando la definición de probabilidad condicional tenemos

$$\mathbb{P}(\text{Prog1} \mid \text{no error}) = \frac{\mathbb{P}(\text{Prog1} \cap \text{no error})}{\mathbb{P}(\text{no error})}$$
$$\mathbb{P}(\text{Progr1} \cap \text{no error}) = 0.15 * 0.95 = 0.1425$$

$\mathbb{P}$(no error)

```
> 0.15*0.95 + 0.20 *0.97 + 0.25 * 0.98 + 0.40 * 0.99

[1] 0.9775
```

$\mathbb{P}(\text{Progr1}|\text{no error}) = 0.1425/0.9775 = 0.1457428.$

$\mathbb{P}(\text{Progr2}|\text{no error})$

```
> 0.20*0.97/0.97775

[1] 0.1984147
```

$\mathbb{P}(\text{Progr3}|\text{no error})$

```
> 0.25*0.98/0.97775

[1] 0.2505753
```

$\mathbb{P}(\text{Progr4}|\text{no error})$

```
> 0.40*0.99/0.97775

[1] 0.4050115
```

2 Eventos independientes

Definición 2.1 Dos eventos A y B son independientes si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Para tres eventos, A, B y C definimos los eventos independientes si

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \\
\mathbb{P}(A \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C) \\
\mathbb{P}(B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C) \text{ y} \\
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)
\end{aligned}$$

Las tres primeras de estas condiciones establecen que los eventos son independientes dos a dos , por lo que llamamos a estos eventos que satisfacen estas tres condiciones como `independientes por pares`. El ejemplo siguiente mostrará que los eventos que satisfacen estas tres condiciones pueden no satisfacer la cuarta condición, de modo que la independencia por pares no determina la independencia.

Para un número finito de eventos $A_1, A_2, \ldots, A_n$ son independientes si

$$\mathbb{P}(A_1 \cap \cdots A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$$

y son independientes por pares si cada par $A_i, A_j, i \neq j$ son independientes.

Si generalizamos un poco, dado un conjunto de eventos $\{A_i : i \in I\}$ estos son independientes si

$$\mathbb{P}\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} \mathbb{P}(A_i)$$

para cada subconjunto J de I.

Existe cierta confusión entre eventos independientes y eventos mutuamente exclusivos. A menudo se habla de estos como `no tener ning?n efecto sobre el otro`, pero eso no es una caracterización precisa en ambos casos. Debes tener en cuenta que aunque los eventos mutuamente exclusivos no pueden ocurrir juntos, los eventos independientes deben ser capaces de ocurrir juntos.

Supongamos que ni $\mathbb{P}(A)$ ni $\mathbb{P}(B)$ es 0 y que A y B son mutualmente exclusivos. Entonces $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$. Así A y B no son independientes. Esto es equivalente a la declaración que si A y B no pueden ser mutualmente exclusivos.

Ejemplo 2.1 Una moneda es lanzada 4 veces. Consideremos los eventos A donde la primera moneda muestra cara; B la tercera moneda muestra sello; y C hay un número igual de caras y sellos. ¿ Esos son eventos independientes ?.

Supongamos que el espacio muestral consiste de 16 puntos que muestran los lanzamientos. El espacio muestral, indicando los eventos que ocurren en cada punto, es como sigue

Puntos	Eventos
C C C C	A
C C C S	A
C C S C	A,B
S C C C	
C S C C	A
C C S S	A,B ,C
C S C S	A,C
S C C S	C
S C S C	B,C
C S S C	A,B,C
S S C C	C
S S S C	B
S S C S	
S C S S	B
C S S S	A,B
S S S S	B

Entonces $\mathbb{P}(A) = 1/2$ y $\mathbb{P}(B) = 1/2$, mientras que C consiste de 6 puntos con exactamente dos caras y dos sellos. Así $P(C) = 6/16 = 3/8$. Ahora $\mathbb{P}(A \cap B) = \dfrac{4}{16} = \dfrac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B); \mathbb{P}(A \cap C) = \dfrac{3}{16} = \mathbb{P}(A) \cdot \mathbb{P}(C)$ y $\mathbb{P}(B \cap C) = \dfrac{3}{16} = \mathbb{P}(B) \cdot \mathbb{P}(C)$. Por tanto los eventos A, B y C son independientes por pares.

El evento $A \cap B \cap C$ consiste de dos puntos $CSSC$ y $CCSS$ con probabilidad $2/16 = 1/8$. Así $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(B)$ y así A, B y C no son independientes.

La siguiente definición generaliza la independencia de una colección arbitraria de eventos, indexados por un conjunto (potencialmente infinito) Θ.

Definición 2.2 Múltiples eventos $A_\theta, \theta \in \Theta$ son independientes por pares si cada par de eventos es independiente. Múltiples eventos $A_\theta, \theta \in \Theta$ son independientes si para cada $k > 0$ y para cada k-subconjunto de distintos eventos $A_{\theta_1}, \ldots, A_{\theta_k}$, tenemos

$$\mathbb{P}(A_{\theta_1} \cap \ldots \cap A_{\theta_k}) = \mathbb{P}(A_{\theta_1}) \cdots \mathbb{P}(A_\theta)$$

La independencia puede surgir de dos maneras distintas. Por ejemplo al lanzar una moneda dos veces, asumimos que los lanzamientos son indepedientes, lo que se `traduce`, en que las monedas `no tienen memoria del primer lanzamiento`. En otro caso, derivamos la independencia verificando $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Por ejemplo, en el lanzamiento de un dado, sea $A = \{2, 4, 6\}$ y sea $B = \{1, 2, 3, 4\}$, entonces

$A \cap B = \{2,4\}, \mathbb{P}(A \cap B) = 2/6 = \mathbb{P}(A)\mathbb{P}(B) = (1/2) \cdot (2/3)$ y así A y B son independientes. En este casono asumimos que A y B son independientes.

Supongamos que A y B son eventos disjuntos, cada uno con probabilidad positiva. ¿ Pueden ser independientes? No. Esto se debe a que $\mathbb{P}(A)\mathbb{P}(B) > 0$ y $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$. Excepto en este caso especial, no hay manera de revisar la independencia mirando los conjuntos en un diagrama de Venn .

Ejemplo 2.2 Lanzamos una moneda 10 veces. Sea el evento $A =$ `al menos una cara`. Sea T_j el evento que salga sello en el j-ésimo lanzamiento. Entonces

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^C) \\
&= 1 - \mathbb{P}(\text{todos los sellos}) \\
&= 1 - \mathbb{P}(T_1 T_2 \cdots T_{10}) \\
&= 1 - \mathbb{P}(T_1)\mathbb{P}(T_2) \cdots \mathbb{P}(T_{10}) \text{ usando independencia} \\
&= 1 - \left(\frac{1}{2}\right)^{10} \approx .999.
\end{aligned}$$

En R

```
> round (1 -(1/2)^{10}, digits = 5)

[1] 0.99902
```

Ejemplo 2.3 Sean dos personas que se turnan para encestar una pelota de baloncesto. La persona 1 tiene éxito con probabilidad 1/3, la persona 2 tiene éxito con probabilidad 1/4. ¿ Cuál es la probabilidad de que la persona 1 tenga éxito antes de la persona 2?. Sea E el evento de interés. Sea A_j el evento en que el primer éxito es de la persona 1 y que se produce en el número de lanzamiento j. Los conjuntos A_i son disjuntos y forman una partición para E, así

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

Ahora $\mathbb{P}(A_1) = 1/3$. A_2 ocurre si tenemos la secuencia persona 1 pierde, persona 2 pierde y la persona 1 tiene éxito. Luego $\mathbb{P}(A_2) = (2/3)(3/4)(1/3) = (1/2)(1/3)$. Siguiendo esa lógica tenemos que $\mathbb{P}(A_j) = (1/2)^{j-1}(1/3)$. Así

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \frac{1}{3}\left(\frac{1}{2}\right)^{j-1} = \frac{1}{3}\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{j-1} = \frac{2}{3}.$$

Proposición 2.1 Si A y B son eventos independientes, entonces $\mathbb{P}(A|B) = \mathbb{P}(A)$. También para algún par de eventos A y B

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

Ejemplo 2.4 Se lanza dos cartas de una baraja, sin reemplazo. Sea A el evento de que el primera carta es el As de Tréboles y sea B el evento de que la segunda corta es la Reina de Diamantes. Entonces $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = (1/52) \times (1/51)$.

```
> (1/51) * (1/52)

[1] 0.0003770739
```

Proposición 2.2 Si A, B son independientes, entonces son independientes los eventos A^c, B, los eventos A, B^c y los eventos A^c, B^c.

3 Regla de Bayes

3.1 Ley de la probabilidad total

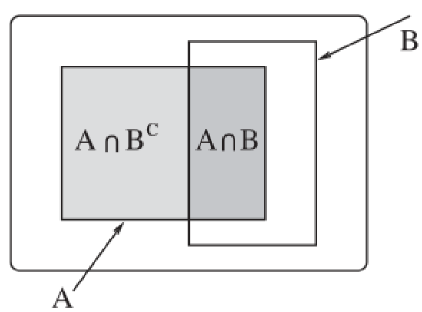
Sea A un evento, entonces se sabe que la intersección de A y el evento universal Ω es A. Se sabe además que B y su complemento B^c constituye una partición. Así

$$A = A \cap \Omega \;\text{ y }\; B \cup B^c = \Omega$$

Sustituyendo el segundo resultado en el primero en la ecuación anterior y aplicando la Ley de Morgan, tenemos

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c). \tag{1}$$

Los eventos $(A \cap B)$ y $(A \cap B^c)$ son mutualmente exclusivos y de acuerdo al siguiente gráfico, se muestra que B y B^c no pueden tener salidas en común, la intersección de A y B no puede tener salidas en común con la intersección de A y B^c



Usando el hecho que $(A \cap B)$ y $(A \cap B^c)$ son mutualmente exclusivos y por propiedad de la función probabilidad, se tiene

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$

Esto significa que para evaluar la probabilidad de un evento A, es suficiente encontrar las probabilidades de la intersección de A y B y A y B^c y sumarlos.

En general sean n eventos $B_i, i = 1, 2, \ldots, n$, una partición del espacio muestral Ω. Entonces para algún evento A, podemos escribir

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i), \quad n \geq 1$$

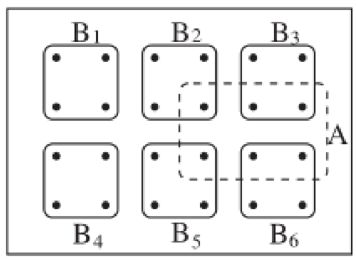
Esta es la ley de la probabilidad total. En efecto, los conjuntos $A \cap B_i, i = 1, 2, \ldots, n$ son mutualmente exclusivos (desde que los B_i lo son) y el hecho que $B_i, i = 1, 2, \ldots$ es una partición de Ω implica que

$$A = \bigcup_{i=1}^{n} A \cap B_i, \; n \geq 1,$$

y por propiedad de la función probabilidad (tercer axioma), se tiene

$$\mathbb{P}(A) = \mathbb{P}\left\{\bigcup_{i=1}^{n} A \cap B_i\right\} = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i).$$

Ejemplo 3.1 Consideramos la siguiente figura que muestra una partición de un espacio muestral conteniendo 24 equiprobables salidas en seis eventos B_1 hasta B_6.



Se sigue entonces que la probabilidad del evento A es igual a $1/4$ ya que contiene seis de los puntos de muestreo. Debido a que los eventos B_i constituyen una partición, cada punto de A está en uno y sólo uno de los eventos B_i y la probabilidad del evento A se pueden encontrar sumando las probabilidades de los eventos $A \cap B_i$ para $i = 1, 2, \ldots, 6$.

Para este ejemplo particular se puede ver que estas seis probabilidades están dadas por $0, 1/24, 1/12, 0, 1/24$ y $1/12$ que cuando se suman juntas da $1/4$.

La Ley de la probabilidad total, es frecuentemente presentado en otro contexto, uno que involucra la probabilidad condicional

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

lo que significa que podemos encontrar $\mathbb{P}(A)$ al encontrar primero la probabilidad de A dado B_i para todo i y luego calcular su promedio ponderado.

Ejemplo 3.2 Mañana habrá lluvia o nieve, pero no ambos, la probabilidad de lluvia es $2/5$ Y la probabilidad de nieve es $3/5$. Si llueve, la probabilidad de que llegue tarde a mi conferencia es $1/5$ mientras que la probabilidad correspondiente en el caso de nieve es $3/5$. ¿Cuál es la probabilidad de que llegue tarde?

Sea A el evento en el que se llega tarde y B sea el evento que llueva. El par B y B^c es una partición del espacio muestral (ya que exactamente uno de ellos debe ocurrir). Por la Ley de la probabilidad total

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \\ &= \frac{1}{5} \cdot \frac{2}{5} + \frac{3}{5} \cdot \frac{3}{5} = \frac{11}{25}.\end{aligned}$$

Ejemplo 3.3 Supongamos que tres cajas contienen bolas blancas y negras. La primera caja contiene 12 bolas blancas y tres negras, la segunda contiene cuatro bolas blancas y 16 negras y la tercera contiene seis bolas blancas y cuatro negras. Se selecciona una caja y se elige una sola bola. La elección de la caja se hace de acuerdo al lanzamiento de un dado. Si el número de puntos en el dado es 1, se selecciona la primera caja, si el número de puntos es 2 ó 3 se elige la segunda caja, de lo contrario (el número de puntos es igual a 4, 5 ó 6) se elige la tercera caja. Supongamos que queremos encontrar $\mathbb{P}(A)$ donde A es el evento en la que una bola blanca es extraida.

En este caso basaremos la partición en las tres cajas. Especificamente, sea $B_i, i = 1,2,3$ el evento en la que la caja i es escogida. Entonces $\mathbb{P}(B_1) = 1/6, \mathbb{P}(B_2) = 2/6$ y $\mathbb{P}(B_3) = 3/6$. Aplicando la Ley de la probabilidad total, tenemos

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3),$$

que es fácil calcular, usando

$$\mathbb{P}(A|B_1) = 12/15, \;\; \mathbb{P}(A|B_2) = 4/20, \;\; \mathbb{P}(A|B_3) = 6/10.$$

así

$$\mathbb{P}(A) = 12/15 \times 1/6 + 4/20 \times 2/6 + 6/10 \times 3/6 = 1/2.$$

Ejemplo 3.4 Un mensaje de correo electrónico puede viajar a través de una de las tres rutas de un servidor. La probabilidad de transmisión de error en cada uno de los servidores y la proporción de mensajes que viajan en cada ruta se muestran en la siguiente tabla. Suponga que los servidores son independientes.

	% mensajes	% errores
Servidor 1	40	1
Servidor 2	25	2
Servidor 3	35	1.5

Determina el porcentaje de mensajes que contienen error

```
> # De la probabilidad total
> 0.4 * 0.01 + 0.25*0.02 + 0.35*0.15

[1] 0.0615
```

3.2 Regla de Bayes

Con frecuencia ocurre que se nos dice que ha ocurrido un cierto evento A y nos gustaría saber cuáles de los eventos mutuamente excluyentes y colectivamente exhaustivos B_j han ocurrido, al menos probabilísticamente. En otras palabras, nos gustaría conocer $\mathbb{P}(B_j|A)$ para cualquier j. Consideremos algunos ejemplos.

En un escenario, se nos puede decir que entre una cierta población hay quienes tienen una enfermedad específica y aquellos que no tienen esa enfermedad. Esto proporciona una partición del espacio muestral (la población) en dos conjuntos disjuntos. Se puede realizar un cierto test, no totalmente fiable en pacientes, con el objeto de detectar la presencia de esta enfermedad.

Si se conoce la proporción de enfermos respecto de los pacientes libres de enfermedad y la fiabilidad del procedimiento de prueba, entonces teniendo en cuenta que un paciente es declarado libre de la enfermedad por este test, queremos saber la probabilidad de que el paciente tiene en realidad la enfermedad (la probabilidad de que el paciente caiga en el primero (o segundo) de los dos conjuntos disjuntos).

El mismo escenario puede obtenerse sustituyendo chips de circuitos integrados por la población y particionándolos en chips defectuosos y no defectuosos, junto con un probador que a veces puede declarar un chip defectuoso bueno y viceversa. Dado que un chip se declara defectuoso por el probabdor, queremos conocer la probabilidad de que sea defectuoso. La transmisión de datos a través de un canal de comunicación sujeto a ruido es todavía un tercer ejemplo. En este caso, la partición es la información que se envía (normalmente 0 y 1 s) y el ruido en el canal puede o no alterar los datos. Escenarios como estos se responden mejor, con la Regla de Bayes.

Obtenemos la regla de Bayes de resultados anteriores sobre la probabilidad condicional y el teorema de la probabilidad total. Así tenemos

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Aunque parezca que esto complica las cosas, lo que estamos haciendo es dividir el problema en piezas más simples. Esto se hace obvio en el ejemplo siguiente donde elegimos un espacio de muestra dividido en tres eventos.

Ejemplo 3.5 Considere un profesor universitario que observa a los estudiantes que entran en su oficina con preguntas. Este profesor determina que el 60% de los estudiantes son estudiantes de BSc, mientras que el 30% son estudiantes de MS, 10% son estudiantes de doctorado. El profesor también indica que puede manejar las preguntas del 80% de los estudiantes BSc en menos de cinco minutos, mientras que las preguntas de 50% de los estudiantes de MS y el 40% de los estudiantes de doctorado en cinco minutos o menos.

El próximo estudiante en ingresar a la oficina del profesor sólo necesitaba dos minutos del tiempo del profesor. ¿Cuál es la probabilidad de que el estudiante sea un estudiante de doctorado?.

Para responder, esta pregunta, sean los eventos $B_i, i = 1, 2, 3$ `el estudiante es un BCs, MS y doctorado` respectivamente y sea el evento A el evento `el estudiante requiere cinco minutos o menos`.

Desde la ley de probabilidad total, tenemos

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3) \\ &= 0.8 \times 0.6 + 0.5 \times 0.3 + 0.4 \times 0.1 = 0.6700\end{aligned}$$

Este cálculo nos da el denominador para la inserción en la regla de Bayes. Esto nos dice que aproximadamente dos tercios de todas las preguntas de los estudiantes se pueden manejar en cinco minutos o menos.

Calculemos $\mathbb{P}(B_3|A)$ usando la regla de Bayes, tenemos

$$\mathbb{P}(B_3|A) = \frac{\mathbb{P}(A|B_3)\mathbb{P}(B_3)}{\mathbb{P}(A)} = \frac{0.4 \times 0.1}{0.6700} = 0.0597$$

O alrededor del 6% y es la respuesta que buscamos.

El punto crítico en la respuesta a preguntas como estas es determinar qué conjunto de eventos constituye la partición del espacio muestral. Las pistas se encuentran generalmente en la pregunta planteada.

Si recordamos que se nos pide calcular $\mathbb{P}(B_j|A)$ y relacionar esto con las palabras en la pregunta, entonces se hace evidente que las palabras `era un estudiante de doctorado` sugiere una partición basada en el estatus del estudiante y el evento A, la información que se nos da, se refiere al tiempo que toma el estudiante.

La clave está en entender que se nos da $\mathbb{P}(A|B_j)$ y se nos pide encontrar $\mathbb{P}(B_j|A)$. En esta forma simple, la ley de Bayes se escribe como

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

Ejemplo 3.6 Un experimentador tienes dos urnas A y B a su disposición. La urna A (B) tiene ocho (diez) canicas verdes y doce(ocho) canicas azules, todas del mismo tamaño y peso. El experimentador selecciona una urna con igual probabilidad y escoge una canica al azar. Se anuncia que la canica seleccionada es azul. Encontremos la probabilidad de que esta canica provenga de la urna B.

Definamos los eventos

A_i: La urna i es seleccionada, $i = 1, 2$.
B : La canica escogida desde la urna seleccionada es azul.

Entonces $\mathbb{P}(A_i) = \dfrac{1}{2}$ para $i = 1, 2$, mientras que $\mathbb{P}(B|A_1) = \dfrac{12}{20}$ y $\mathbb{P}(B|A_2) = \dfrac{8}{18}$.

Ahora apliquemos el teorema de Bayes,

$$\mathbb{P}(A_2|B) = \mathbb{P}(A_2)P(B|A_2)/\{\mathbb{P}(A_1)P(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2)\}$$

Reemplazando en R

```
> (1/2)*(8/18)/((1/2)*(12/20) + (1/2)*(8/18))

[1] 0.4255319
```

4 Muestreo Aleatorio

Uno de los conceptos importantes en probabilidades y estadística, es el de *muestra aleatoria*, que se presenta en diversos aspectos, como la extracción de una carta desde una baraja. En R se puede simular esas situaciones con la función `sample`.

Por ejemplo si se quiere obtener 5 números de manera aleatoria, de un conjunto `1:40`, entonces podemos escribir

```
> sample(1:40, 5)

[1] 16 10 19  6  8
```

Se debe notar que el comportamiento predeterminado de `sample` es el de *muestreo sin reemplazo*. Es decir, las muestras no contienen el mismo número dos veces, y el tamaño, obviamente, no puede ser más grande que la longitud del vector a muestrear.

Si se desea que el muestreo sea con reemplazo, entonces debemos agregar el argumento `replace = TRUE`.

El muestreo con reemplazo es adecuado para el modelado de lanzamientos de monedas o lanzamientos de un dado. Así, para simular 12 veces el lanzamiento de una moneda podríamos escribir

```
> sample(c("H","T"), 12, replace=T)

 [1] "T" "T" "H" "T" "T" "H" "H" "T" "T" "T" "H" "H"
```

En el lanzamientos de monedas, la probabilidad de caras debe ser igual a la probabilidad de sellos, pero la idea de un evento al azar no se limita a casos simétricos. Se podría igualmente aplicarse a otros casos, como el resultado exitoso de un procedimiento quirúrgico.

Podemos simular datos con probabilidades no iguales para los resultados (por ejemplo, un 90% de probabilidades de éxito) utilizando el argumento `prob` a `sample`

```
> sample(c("exito", "fallo"), 10, replace=T, prob=c(0.9, 0.1))

 [1] "fallo" "exito" "exito" "exito" "exito" "exito" "fallo" "exito" "exito" "exito"
```

Ejemplo 4.1 Simular el lanzamiento de una moneda puede ser hecho usando la función *rbinom* en lugar de *sample*. En efecto

```
> rbinom(10, 1, .5)

 [1] 1 1 0 1 0 0 1 0 1 1
```

o de la siguiente manera

```
> ifelse(rbinom(10, 1, .5) == 1, "H", "T")

 [1] "T" "H" "T" "T" "T" "H" "H" "T" "T" "H"
```

```
> c("H", "T")[1 + rbinom(10, 1, .5)]

 [1] "H" "H" "H" "H" "T" "T" "T" "T" "H" "H"
```

5 Referencias

1. Introduction to Probability and Statistics, Lecture 4 Bayes Law KC Border 2016.

2. Probability (chapter 1) All of Statistics A concise course in Statistical Inference Larry Wassermann Springer 2004.