El Teorema de Bayes con Diagramas de Venn

Julio Sergio Santana

2019-07-04

Índice general

Prólogo

El teorema de Bayes es un tema central de la estadística. No obstante, las explicaciones del tema o bien son muy *matemáticas*, o no son lo suficientemente claras para ahondar en el objetivo que el teorema persigue. El breve texto que se presenta a continuación pretende subsanar dicha carencia por medio de la elaboración de una explicación gráfica basada en los diagramas de Venn.

Capítulo 1

Teorema de Bayes Explicado

Antes de proceder a la explicación del teorema, se enunciará éste de una manera plana.

1.1. El teorema de Bayes

Teorema 1.1 (Teorema de Bayes)**.** *Dados los conjuntos (eventos) $B_1, B_2, ...B_k$, disjuntos, y los conjuntos S, que representa el universo, y A, tales que,*

$$S = \bigcup_{i=1}^{k} B_i$$
$$A \subseteq S$$

se cumple que,

$$P\left(B_j|A\right) = \frac{P\left(B_j\right) \cdot P\left(A|B_j\right)}{\sum_{i=1}^{k} P\left(B_i\right) \cdot P\left(A|B_i\right)} \tag{1.1}$$

1.2. Explicación gráfica del teorema

1.2.1. Situación inicial

La Figura 1.1 muestra, en un diagrama de Venn, más o menos la situación inicial planteada en el Teorema 1.1. Aquí se puede ver, por ejemplo, que todos los con-

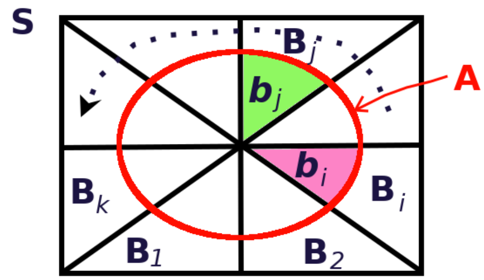

Figura 1.1: Diagrama de Venn para el teorema de Bayes

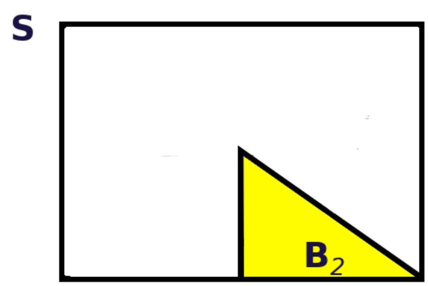

Figura 1.2: Relación de áreas: ¿qué porción de S ocupa B_2?

juntos disjuntos, B_i, representados por los triángulos de la figura, componen al conjunto S, representado por el rectángulo mayor de la figura, o que el conjunto A, representado por el óvalo rojo, está contenido en el conjunto S.

Si se hiciera un símil entre las relaciones de áreas de la Figura 1.1, con las probabilidades de los eventos, por ejemplo, la probabilidad del evento B_2, estaría dada por la relación de áreas que responde a la pregunta: *¿qué porción de S ocupa B_2?*, y que se muestra en la siguiente fórmula y en la Figura 1.2:

$$P\left(B_2\right) = \frac{area\left(B_2\right)}{area\left(S\right)}$$

Con el fin de simplificar, la fórmula anterior la denotaremos sólo con

$$P\left(B_2\right) = \frac{B_2}{S} \tag{1.2}$$

Estrictamente hablando, el lado izquierdo de la igualdad en la Ecuación (1.2), debería haberse escrito como $P\left(B_2|S\right)$. Sin embargo, como S representa el uni-

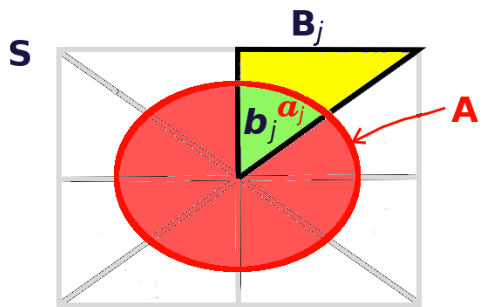

Figura 1.3: Relación de áreas: ¿qué porción de A o de B_j ocupa $b_j(=a_j)$?

verso o espacio total de muestreo, se obvia en la fórmula, y se escribe simplemente como $P\left(B_2\right)$.

En el mismo tenor, y tomando como referencia la Figura 1.3, si $(B_j \cap A) = b_j = a_j$, hay dos preguntas interesantes aquí, a saber:

1. *¿Qué porción de A ocupa b_j?*, o lo que es lo mismo, *¿Cuánto de A está ocupando B_j?*, y en términos de probabilidad:

$$P\left(B_j|A\right) = b_j/A \tag{1.3}$$

2. *¿Qué porción de B_j ocupa a_j?*, o lo que es lo mismo, *¿Cuánto de B_j está ocupando A?*, y en términos de probabilidad:

$$P\left(A|B_j\right) = a_j/B_j \tag{1.4}$$

1.2.2. El teorema de la probabilidad total

De acuerdo con la Figura 1.1, y como $A \subseteq S$, se cumple que:

$$A = A \cap S \tag{1.5}$$

Como $S = \bigcup_{i=1}^{k} B_i$, se puede sustituir en la Ecuación (1.5), con lo que se obtiene:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup ... (A \cap B_k) \tag{1.6}$$

Si se toma la probabilidad en ambos lados de la ecuación y como los B_i son disjuntos, se obtiene:

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k) \\ &= \sum_{i=1}^{k} P(A \cap B_i) \end{aligned} \tag{1.7}$$

El teorema del producto de probabilidades de eventos dependientes (Walpole et al., 2012, pp. 66) establece que

$$\begin{aligned} P(X \cap Y) &= P(X) \cdot P(Y|X) \\ &= P(Y) \cdot P(X|Y) \end{aligned} \tag{1.8}$$

Si se aplica esto a la Ecuación (1.7), se tiene el

Teorema 1.2 (Teorema de la probabilidad total)**.** *Dados los conjuntos (eventos) $B_1, B_2, ...B_k$, disjuntos, y los conjuntos S, que representa el universo, y A, tales que,*

$$\begin{aligned} &S = \bigcup_{i=1}^{k} B_i \\ &A \subseteq S \end{aligned}$$

se cumple que,

$$P(A) = \sum_{i=1}^{k} P(B_i) \cdot P(A|B_i) \tag{1.9}$$

Gráficamente, si se toma como base la Ecuación (1.4), que representa la *relación de áreas* que se muestra en la Figura 1.3, y se introduce en la Equación (1.9), se tendría que:

$$\begin{aligned} P(A) &= \sum_{i=1}^{k} \frac{B_i}{S} \frac{a_i}{B_i} \\ &= \frac{1}{S} \sum_{i=1}^{k} a_i = \frac{A}{S} \end{aligned} \tag{1.10}$$

1.2.3. El teorema de Bayes: deducción y explicación analítica

Para deducir el teorema de Bayes, enunciado en el Teorema 1.1 al principio de este capítulo, se parte del teorema del producto de probabilidades de eventos dependientes, dado en la Ecuación (1.8). En este caso, se aplica a los eventos A y uno, cualquiera, de los B_j, así:

$$P(A) \cdot P(B_j|A) = P(A \cap B_j) = P(B_j) \cdot P(A|B_j) \tag{1.11}$$

Si se despeja, se obtiene:

$$P(B_j|A) = \frac{P(B_j) \cdot P(A|B_j)}{P(A)} \tag{1.12}$$

Finalmente se sustituye $P(A)$, haciendo uso del Teorema de la probabilidad total (1.2), dado en la Ecuación (1.9), y se llega al teorema de Bayes:

$$P(B_j|A) = \frac{P(B_j) \cdot P(A|B_j)}{\sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)} \tag{1.13}$$

Para explicar el significado del teorema, se recurrirá a la Figura 1.4. La parte (1) de la figura representa el lado izquierdo de la igualdad en la Ecuación (1.13), las partes (2) y (3), representan el numerador del cociente en el lado derecho de igualdad en la ecuación, y la parte (4) representa el denominador.

A continuación se explica cada una de las partes de la Figura 1.4.

1.2.3.1. Parte (1): Lado izquierdo y *objetivo* del teorema

El lado izquierdo de la igualdad en la Ecuación (1.13), establece el *objetivo*, por así decirlo, del teorema; esto es, lo que se quiere obtener. En este caso es, $P(B_j|A)$, lo que gráficamente, como se muestra en la parte (1) de Figura 1.4, es la porción que B_j ocupa del área A. En términos estadísticos, es la probabilidad de que ocurra B_j dado que ocurrió A. Gráficamente, y en términos de relaciones de áreas esto es:

$$P(B_j|A) = \frac{b_j}{A} \tag{1.14}$$

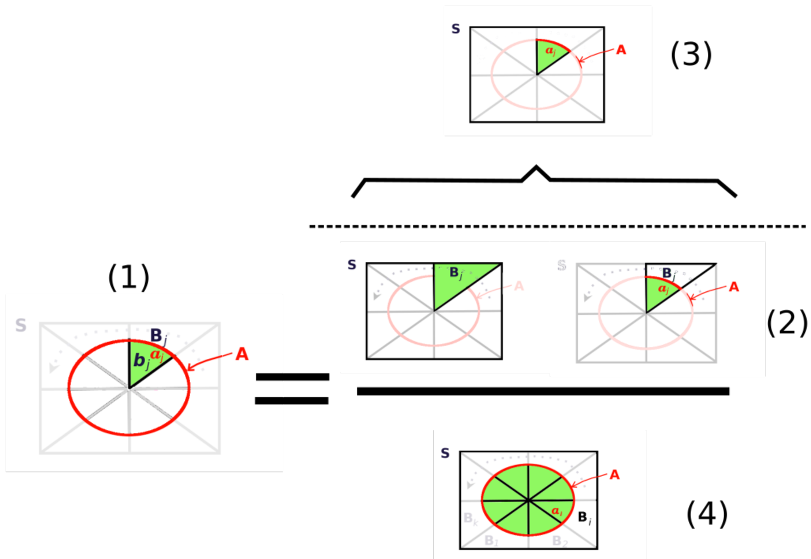

Figura 1.4: Explicación gráfica del teorema de Bayes

1.2.3.2. Partes (2) y (4): Lado derecho, el *método* del teorema

El lado derecho de la igualdad en la Ecuación (1.13), establece el *método*, esto es, la forma o manera de obtener lo que se ha establecido en el *objetivo*. En este caso, se trata de un cociente compuesto de dos partes, el numerador y el denominador.

Es importante notar que, en la fórmula del teorema, dada en la Ecuación (1.13), la estructura del numerador y el denominador es semejante: incluye un producto del tipo $P(B_r) \cdot P(A|B_r)$, donde, ya sea, $r = j$, o $r = i$, salvo que en el caso del numerador sólo se trata de uno de estos productos, mientras que en el caso del denominador se trata de una suma de los productos, de tal manera que, lo que se diga acerca del producto registrado en el numerador, aplica a cada uno de los productos de la suma registrada en el denominador.

1.2.3.3. Parte (2): El numerador

El numerador es un producto compuesto de dos factores: $P(B_j)$ y $P(A|B_j)$, que corresponden respectivamente a cada una de las dos gráficas en la parte (2) de la Figura 1.4. El primer factor, y la correspondiente primera gráfica, representan la porción que ocupa B_j del *universo*, S, esto es, la relación de áreas,

$$P\left(B_j\right)=\frac{B_j}{S} \tag{1.15}$$

El segundo factor, y la correspondiente segunda gráfica, representan la porción que ocupa A de B_j, esto es, la relación de áreas,

$$P\left(A|B_j\right)=\frac{a_j}{B_j} \tag{1.16}$$

1.2.3.4. Parte (3): El numerador

Si se procede a desarrollar el producto de los lados derechos de la Ecuaciones (1.15) y (1.16), se obtiene:

$$\frac{B_j}{S}\cdot\frac{a_j}{B_j}=\frac{a_j}{S} \tag{1.17}$$

Que es la porción que ocupa a_j, resultado de la intersección entre A y B_j, del *universo*, S, y que es precisamente lo que registra la gráfica de la parte (3) de la Figura 1.4.

1.2.3.5. Parte (4): El denominador

Como se ha mencionado en la Sección 1.2.3.2, el denominador es la sumatoria de productos del tipo $P\left(B_r\right)\cdot P\left(A|B_r\right)$, y, por consiguiente, para cada uno de éstos aplica lo desarrollado en las dos secciones anteriores (1.2.3.3 y 1.2.3.4). Esto es, para cada elemento del tipo B_i, se tendría un resultado similar al expresado por la gráfica de la parte (3) de la Figura 1.4. Al sumar todos esos resultados, se obtiene justamente lo expresado por la gŕafica de la parte (4) de la Figura 1.4, que representa la reconstrucción del área A en relación con el área que representa el *universo*, S. Vale decir que esta reconstrucción es propiamente lo que establece el Teorema 1.2 de la probabilidad total. Esta sumatoria resulta entonces en

$$\begin{aligned}\sum_{i=1}^{k}P\left(B_i\right)\cdot P\left(A|B_i\right)&=\frac{A}{S}\\&=P\left(A|S\right)=P\left(A\right)\end{aligned} \tag{1.18}$$

1.2.3.6. El cociente

Al dividir el numerador establecido en la Ecuación (1.17), entre el denominador, tal como se establece en la Ecuación (1.18), se llega al siguiente resultado:

$$\frac{\frac{a_j}{S}}{\frac{A}{S}} = \frac{a_j}{A} \tag{1.19}$$

Este resultado coincide con el *objetivo* establecido en la Sección 1.2.3.1, en la Ecuación (1.14) y con la parte (1) de la Figura 1.4, ya que $a_j = b_j$.

1.2.3.7. Colofón

En la Sección 1.2.3.1 se estableció el *objetivo* del teorema, que es obtener $P(B_j|A)$ y, a partir de ahí, en las siguientes secciones se analizó el método para llegar a ese objetivo que es a través del cálculo de un conjunto de productos del tipo $P(B_i) \cdot P(A|B_i)$. Cabe notar que, uno de estos productos es precisamente el que figura en el numerador del teorema, a saber, $P(B_j) \cdot P(A|B_j)$, que corresponde al evento B_j del *objetivo* del teorema. El denominador, por su parte, es la suma de todos los productos calculados.

Capítulo 2

Ejemplos

A continuación se muestran algunos ejemplos para clarificar el uso del teorema y para mostrar su utilidad. Para resolver los problemas dados en los ejemplos, se usará el lenguaje de programación R (Santana and Mateos Farfán, 2014).

2.1. Ejemplo 1

Una empresa utiliza filtros procedentes de dos marcas de *software* para detectar correo *spam*: la marca **Ms** y la marca **G**. Estadísticamente se ha determinado a nivel mundial que **Ms** clasifica correctamente sólo el 70 % de los correos como *spam* o *no-spam*, mientras que **G** alcanza el 98 %. Sin embargo, como la junta directiva se ha dejado llevar por la marca, ha determinado que 65 % de sus correos sean filtrados por el *software* de **Ms**, mientras que el resto, o sea, el 35 %, sean asignados para ser filtrados por el *software* de la marca **G**.

a. Roberto, el ingeniero de sistemas de la empresa, que es mucho más consciente de la realidad, quiere convencer a la junta directiva de cambiar su política de asignación y para ello se ha propuesto responder a la pregunta: ¿Cuál es la probabilidad de que un correo bien detectado, ya sea como *spam* o *no-spam*, proceda de **Ms**?

Solución

Primeramente se procede a identificar los distintos eventos del problema.

1. A = se ha clasificado correctamente el correo, ya sea como *spam* o como *no-spam*.

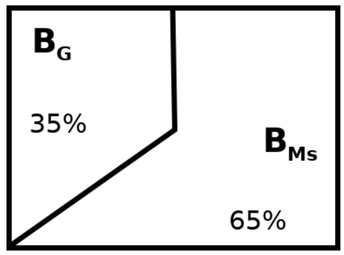

Figura 2.1: Diagrama de Venn que muestra la asignación de correos por marca de software

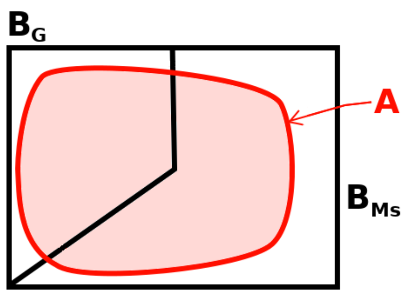

Figura 2.2: Diagrama de Venn que muestra el evento A: el correo ha sido clasificado correctamente

2. B_{Ms} = se ha asignado al *software* de la marca **Ms** para filtrar el correo.
3. B_G = se ha asignado al *software* de la marca **G** para filtrar el correo.

La asignación total del volumen de correos por marca de *software* se muestra en la Figura 2.1.

Por otra parte, el evento A que consiste en haber clasificado correctamente el correo, se muestra en relación con el universo de correos en la Figura 2.2.

Se hará una tabla (*data.frame*) para capturar toda la información del problema:

```
datos <- data.frame(
  #              marca.Ms     marca.G
  P.B       =  c(    0.65   ,    0.35     ),  # P(B)
  P.A_B     =  c(    0.70   ,    0.98     ),  # P(A|B)
  row.names =  c("marca.Ms" , "marca.G"   )
```

Tabla 2.1: Información distribución de correos

	P.B	P.A_B
marca.Ms	0.65	0.70
marca.G	0.35	0.98

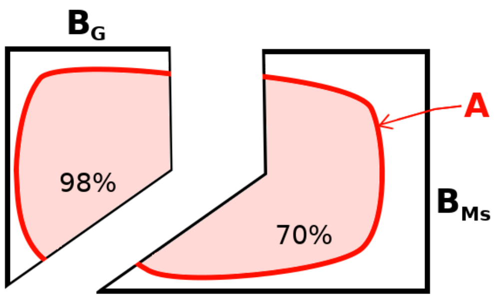

Figura 2.3: Diagrama de Venn que muestra la porción que ocupa A: de cada una de las áreas B

```
)
print(datos)
```

```
##          P.B P.A_B
## marca.Ms 0.65  0.70
## marca.G  0.35  0.98
```

Esta información se muestra en la Tabla 2.1 a continuación.

En la tabla anterior, la columna P.A_B, es $P(A|B)$, o sea la probabilidad de que se haya clasificado bien el correo, dado que la clasificación se hizo con el *software* correspondiente al renglón de la tabla. Esto es, en términos de *relaciones de áreas*, la porción que ocupa A del conjunto B correspondiente. Esta situación se muestra gráficamente en la Figura 2.3

A continuación se procede a aplicar el Teorema 1.2 de la probabilidad total. Esto es multiplicar las columnas "P.B" y "P.A_B" de la Tabla 2.1 y sumar todos los resultados.

```
datos$Prod <- datos$P.B * datos$P.A_B
# La suma de las columnas
sumas <- apply(datos, 2, sum)
# Se agrega el resultado como un renglón al final:
datos["SUMAS", ] <- sumas
```

Tabla 2.2: Productos de probabilidades en el caso de correos

	P.B	P.A_B	Prod
marca.Ms	0.65	0.70	0.455
marca.G	0.35	0.98	0.343
SUMAS	1.00	1.68	0.798

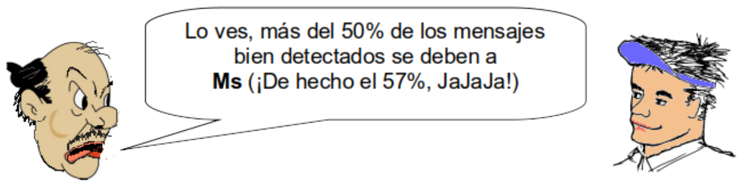

Figura 2.4: Respuesta de los directivos de la empresa

El resultado de este producto se puede ver en la columna **Prod** de la Tabla 2.2

De acuerdo con este resultado, la probabilidad de que, en general, un correo sea clasificado correctamente es del 79.8 %.

Para averiguar la probabilidad de que un correo bien detectado, ya sea como *spam* o *no-spam*, proceda de **Ms**, se aplica el Teorema 1.1 de Bayes. Para ello se divide el producto en el renglón correspondiente a **Ms** entre la probabilidad recien obtenida, así:

```
(Prob.Ms <- datos["marca.Ms","Prod"]/datos["SUMAS","Prod"])
```

```
## [1] 0.5701754
```

Esto es, dado que se clasificó correctamente el correo, la **probabilidad de que para esta operación se haya usado el *software* de Ms** es de 57.02 %.

Después de presentar estos resultados a la junta directiva de la empresa, la respuesta que Roberto recibió no fue muy amable, como se puede ver en la ilustración de la Figura 2.4.

Aunque en un principio la respuesta lo tomó por sorpresa, en vez de desanimarse, Roberto trató de replantear el problema desde una prespectiva diferente, tal como se ilustra en la Figura 2.5.

Esto da lugar a la segunda parte de este problema, cuya descripción se hace en seguida:

b. Determinar la probabilidad que que un correo mal clasificado, ya sea como *spam* o *no-spam*, provenga del *software* provisto por **Ms**.

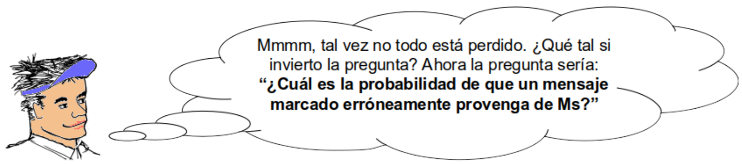

Figura 2.5: El problema visto desde otra perspectiva

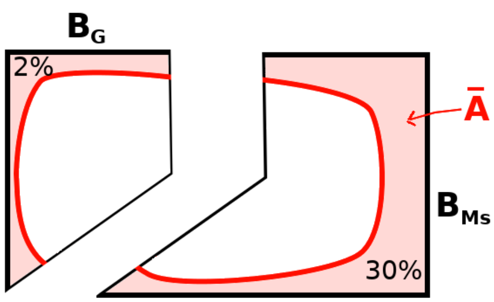

Figura 2.6: Diagrama de Venn que muestra la porción que ocupa $\bar{A}$: de cada una de las áreas B

Con el propósito de responder a esta pregunta, debe notarse que ahora el tema ya no es el evento A sino su complemento, $\bar{A}$. Para resolver el problema, solamente es necesario modificar la columna “P.A_B”, de la tabla anterior, de acuerdo con los nuevos datos mostrados en la Figura 2.6 y proceder a los cálculos correspondientes.

```
#                  marca.Ms      marca.G  SUMA
datos$P.A_B <- c(    0.30    ,     0.02,     NA ) # P(A|B)
# Se corrigen los productos
datos$Prod <- datos$P.B * datos$P.A_B

# Se agregan las sumas corregidas como el renglón final:
datos["SUMAS", ] <- apply(datos[-3, ], 2, sum)
```

En la Tabla 2.3 se pueden apreciar los resultados de estas operaciones.

Finalmente, se aplica el Teorema 1.1 de Bayes. Para ello se divide el producto en el renglón correspondiente a **Ms** entre la probabilidad total, registrada el renglón “SUMAS” y columna “Prod”, así:

```
(Prob.Ms <- datos["marca.Ms","Prod"]/datos["SUMAS","Prod"])
```

```
## [1] 0.9653465
```

Tabla 2.3: Productos de probabilidades en el caso de correos mal clasificados

	P.B	P.A_B	Prod
marca.Ms	0.65	0.30	0.195
marca.G	0.35	0.02	0.007
SUMAS	1.00	0.32	0.202

Esto es, dado que se clasificó incorrectamente el correo, la **probabilidad de que para esta operación se haya usado el *software* de Ms** es de 96.53 %. Con este contundente resultado, Roberto pudo convencer a la junta directiva de cambiar las proporciones de asignación de correos a las distintas marcas de *software*.

¿**Qué hubiera pasado si las estadísticas del problema fueran diferentes**? Intrigado por esta pregunta, Roberto decidió resolver el problema de una manera más general. A continuación se muestra una aplicación que permite jugar con los valores de las distintas variables del problema.

2.2. Ejemplo 2

Por razones de pagos en publicidad, un piloto de autos usa un Corvette en el 50 % de las carreras en las que participa, un Jaguar, en el 30 % de esas carreras y un Alfa Romeo, en el 20 % de las mismas. De 25 carreras en las que ha participado

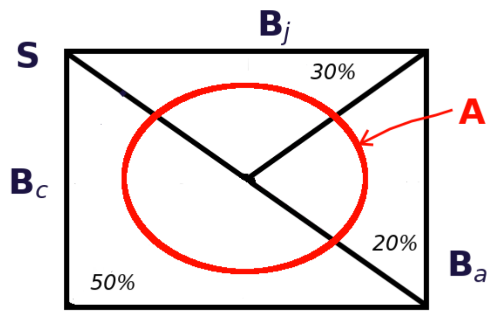

Figura 2.7: Diagrama de Venn para el problema del piloto de autos

con el Corvette, ha ganado 5; de 15, en las que ha participado con el Jaguar, ha ganado 4; y de 10, en las que ha participado con el Alfa Romeo, ha ganado 4.

a. Haciendo uso de esa información para estimar las probabilidades, indique cuál es la probabilidad de que el piloto gane la reciente carrera en la que participará en Le Mans.
b. Suponiendo que llega la notificación de que, en efecto, ganó la carrera, ¿cuál es la probabilidad de que haya manejado el Corvette?

NOTA: Este ejemplo fue tomado de los ejercicios del libro de Miller and Freund (1965) pp. 32-34.

Solución

La Figura 2.7 representa esquemáticamente algunos de los elementos del problema, en esta figura:

1. S representa el conjunto de todas las carreras que ha corrido el piloto.
2. B_c son las carreras en las que ha usado el Corvette.
3. B_j son las carreras en las que ha usado el Jaguar.
4. B_a son las carreras en las que ha usado el Alfa Romeo.
5. A representa el conjunto de carreras en las que ha ganado.

Para resolver el problema se usará el lenguaje de programación R (Santana and Mateos Farfán, 2014). El primer paso es la construcción de una tabla con la información

Tabla 2.4: Información del piloto de autos

	P.B	P.A_B
Corvette	0.5	0.2000000
Jaguar	0.3	0.2666667
Alfa_Romeo	0.2	0.4000000

Tabla 2.5: Productos de probabilidades

	P.B	P.A_B	Prod
Corvette	0.5	0.2000000	0.10
Jaguar	0.3	0.2666667	0.08
Alfa_Romeo	0.2	0.4000000	0.08

```
datos <- data.frame(
  #                 Corvette    Jaguar    Alfa_Romeo
  P.B         =  c(    0.5   ,    0.3   ,    0.2       ), # P(B)
  P.A_B       =  c(   5/25   ,   4/15   ,   4/10       ), # P(A|B)
  row.names   =  c("Corvette", "Jaguar" , "Alfa_Romeo")
)
print(datos)
```

```
##              P.B     P.A_B
## Corvette   0.5 0.2000000
## Jaguar     0.3 0.2666667
## Alfa_Romeo 0.2 0.4000000
```

Esta misma información se puede ver en la Tabla 2.4, donde la columna **P.B** representa la probabilidad del conjunto B en el renglón correspondiente, y la columna **P.A_B** es la probabilidad de que se ganó la carrera dado que se usó el auto en el renglón correspondiente.

La solución del inciso **a** del problema se obtiene por medio de la aplicación del Teorema 1.2 de la probabilidad total. Esto es multiplicar las columnas "P.B" y "P.A_B" de la tabla anterior y sumar todos los resultados.

```
datos$Prod <- datos$P.B * datos$P.A_B
```

El resultado de este producto se puede ver en la columna **Prod** de la Tabla 2.5

Ahora se procede a hacer la suma de los valores en la columna **Prod** de la Tabla 2.5 con

```
(Prob.A <- sum(datos$Prod))
```

```
## [1] 0.26
```

Asi entonces, la **probabilidad de que el piloto gane su carrera** es del 26 %.

Para resolver el inciso **b** del problema se aplica el Teorema 1.1 de Bayes. En este caso es simplemente dividir el producto en el renglón correspondiente al auto “Corvette” entre la probabilidad recien obtenida, así:

```
(Prob.Corvette <- datos["Corvette","Prod"]/Prob.A)
```

```
## [1] 0.3846154
```

Esto es, dado que el piloto en efecto ganó la carrera, la **probabilidad de que haya usado el Corvette** es de 38.46 %.

Bibliografía

Miller, I. and Freund, J. E. (1965). *Probability and statistics for engineers*. Prentice-Hall Mathematics Series. Prentice-Hall. Englewood Cliffs, NJ, U.S.A., 1st edition.

Santana, J. S. and Mateos Farfán, E. (2014). *El arte de programar en R: un lenguaje para la estadística*. Instituto Mexicano de Tecnología del Agua, Jiutepec, Mor., México, 1 edition.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (2012). *Probability & Statistics for Engineers & Scientists*. Prentice Hall / Pearson Education, Inc., Boston, MA, 9th edition. ISBN 10: 0-321-62911-6.