1 El teorema de Bayes

1.1 Ley de la probabilidad total

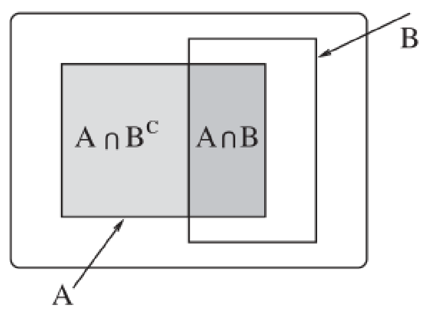
Sea A un evento, entonces se sabe que la intersección de A y el evento universal Ω es A. Se sabe además que B y su complemento B^c constituye una partición. Así

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

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

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

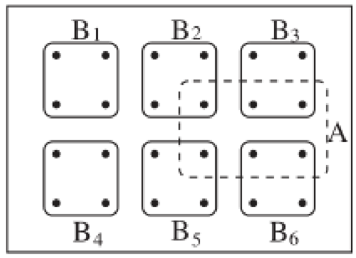
Esta es la ley de la probabilidad total. En efecto, los conjuntos $A \cap B_i, i = 1, 2, \ldots, n$ son mutualmente exclusivos (desde que los B_i lo son) y el hecho que $B_i, i = 1, 2, \ldots$ es una partición de Ω implica que

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

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 1.1 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 1.2 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

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

Ejemplo 1.3 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 1.4 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
```

1.2 Teorema 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 1.5 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

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

Ejemplo 1.6 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:

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

[1] 0.4255319
```

2 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.