Introducción a la probabilidad y estadística CM274

César Lara Avila

9 de agosto de 2017

https://github.com/C-Lara

3. Probabilidad condicional, eventos independientes Regla de Bayes

Probabilidad condicional

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 $\mathbb{P}(\bigcup\limits_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$.

Pero no es cierto en general que $\mathbb{P}(A|B \cup C) = \mathbb{P}(A|B) + P(A|C)$.

Propiedades de la probabilidad condicional

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

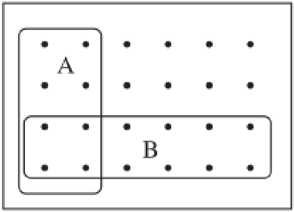
  $$\mathbb{P}_A(E) = \mathbb{P}(E|A)$$

  y todas las propiedades y intuiciones que se aplican a la función probabilidad se aplican a $\mathbb{P}_A$ también.

Ejemplo explicativo

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 contraintuitivo Feller 1968

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

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

Independencia

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

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

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`. 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.

Generalización del concepto

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.

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

Contraejemplo explicativo

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.

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 $\mathbb{P}(A \cap B) = \frac{4}{16} = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B); \mathbb{P}(A \cap C) = \frac{3}{16} = \mathbb{P}(A)\mathbb{P}(C)$ y $\mathbb{P}(B \cap C) = \frac{3}{16} = \mathbb{P}(B)\mathbb{P}(C)$. 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 $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(B)$. Luego los evento A, B y C no son independientes.

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

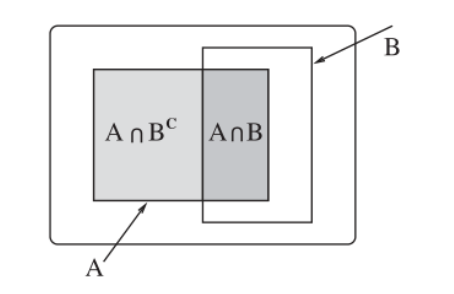
$$A = A \cap \Omega \ \text{ y } \ B \cup B^c = \Omega$$

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

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Ley de la probabilidad total(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,

Ley de la probabilidad total(2)

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.

Ley de la probabilidad total(3)

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

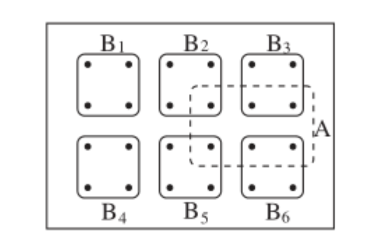
$$A = \bigcup_{i=1}^{n} A \cap B_i, \;\; n \geq 1,$$

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 explicativo

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 $1/4$.

Uso de árboles para organizar el cálculo

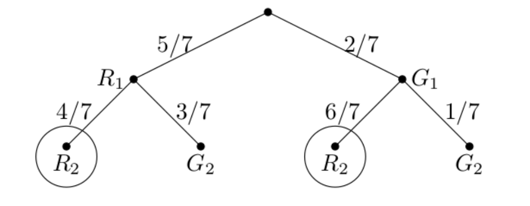
Los árboles son una manera de organizar cálculos con probabilidad condicional y la ley de probabilidad total. Las figuras y los ejemplos muestran lo que queremos decir con un árbol. Al igual que con la regla del producto, la clave es organizar el proceso subyacente en una secuencia de acciones.

Sea el siguiente ejemplo: Una urna contiene 5 pelotas rojas y 2 pelotas gris. Se saca una pelota. Si es gris, se agrega una pelota roja a la urna y si es roja, se agrega una pelota gris a la urna. (La pelota original no se devuelve a la urna.) Luego se saca una segunda pelota. ¿Cuál es la probabilidad de que la segunda pelota sea roja?.

Comenzamos rehaciendo el ejemplo anterior. La secuencia de acciones es: primero se saca la pelota 1 (y añade la bola apropiada a la urna) y luego se saca la pelota 2.

Uso de árboles para organizar el cálculo(1)

Sea el siguiente diagrama de árbol del ejemplo anterior:



Se interpreta este árbol de la siguiente manera. Cada punto se llama nodo. El árbol está organizado por niveles. El nodo superior (nodo raíz) está en el nivel 0. La capa siguiente es el nivel 1 y así sucesivamente. Cada nivel muestra los resultados en una etapa de la extracción. El nivel 1 muestra los posibles resultados de la primera extracción. El nivel 2 muestra los posibles resultados de la segunda extracción a partir de cada nodo en el nivel 1.

Las probabilidades se escriben a lo largo de las ramas. La probabilidad de que R_1 (rojo en la primera extracción) es 5/7. Esto se escribe a lo largo de la rama desde el nodo de raíz a un nodo etiquetada por R_1. En el siguiente nivel ponemos probabilidades condicionales. La probabilidad a lo largo de la rama de R_1 a R_2 es $\mathbb{P}(R_2|R_1) = 4/7$ que representa la probabilidad de ir al nodo R_2 dado que ya se está en R_1.

Uso de árboles para organizar el cálculo(2)

La regla de multiplicación dice que la probabilidad de llegar a cualquier nodo es sólo el producto de las probabilidades a lo largo de la ruta para llegar allí. Por ejemplo, el nodo etiquetado R_2 en la extrema izquierdo representa realmente el evento $R_1 \cap R_2$ porque proviene del nodo R_1. La regla de multiplicación dice:

$$\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_1)\mathbb{P}(R_2|R_1) = \frac{5}{7} \cdot \frac{4}{7},$$

que se está multiplicando exactamente a lo largo de la ruta hacia el nodo.

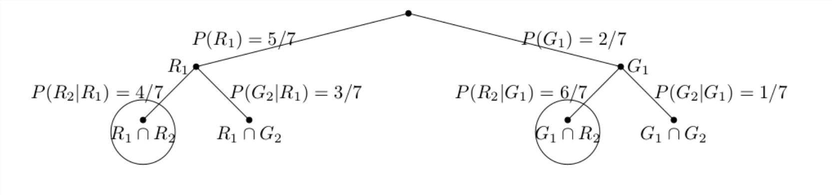
La ley de probabilidad total es simplemente la afirmación de que $\mathbb{P}(R_2)$ es la suma de las probabilidades de todos las rutas que conducen a R_2 (los dos nodos en círculo en la figura). En este caso,

$$\mathbb{P}(R_2) = \frac{5}{7} \times \frac{4}{7} + \frac{2}{7} \times \frac{6}{7} = \frac{32}{49}.$$

Árboles precisos

El árbol anterior implica ciertas simplificaciones. Por ejemplo, el nodo R_2 en el extremo izquierdo representa realmente el evento $R_1 \cap R_2$, ya que termina en la ruta, desde la raíz hasta R_1 hasta R_2.

Aquí está el mismo árbol etiquetado con precisión.



Como se puede ver este árbol es más complicado de hacer y de usar. Usualmente se usa una versión abreviada de los árboles.

El teorema de Bayes

El teorema de Bayes es un pilar de la probabilidad y estadística y es central para el resto de este curso y otros cursos más avanzados . Para dos eventos A y B el teorema de Bayes (también llamado la regla de Bayes y fórmula de Bayes) dice,

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

- La regla de Bayes nos dice cómo invertir las probabilidades condicionales, es decir, encontrar $\mathbb{P}(B|A)$ de $\mathbb{P}(A|B)$.
- En la práctica, $\mathbb{P}(A)$ a menudo se calcula utilizando la ley de probabilidad total.

En efecto, obtenemos la regla de Bayes de resultados anteriores sobre la probabilidad condicional y el teorema de la probabilidad total. Así tenemos

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

Ejemplo explicativo

Considere una prueba de rutina de detección de una enfermedad. Supongamos que la frecuencia de la enfermedad en la poblacin (tasa básica) es del $0,5\,\%$. La prueba es muy precisa con un $5\,\%$ de tasa de falsos positivos y un $10\,\%$ de tasa de falsos negativos. Si se toma la prueba y es positivo. ¿Cuál es la probabilidad de que se tenga la enfermedad?

Haremos el cálculo tres veces: usando árboles, tablas y símbolos. Utilizamos la siguiente notación para los eventos relevantes:

- $D+ =$ 'tienes la enfermedad'.
- $D- =$ 'no tienes la enfermedad'.
- $T+ =$ 'prueba positiva'
- $T- =$ 'prueba negativa'.

Se nos da $\mathbb{P}(D+) = .005$ y por lo tanto $\mathbb{P}(D-) = .995$. Las tasas falsas positivas y falsas negativas son (por definición) probabilidades condicionales.

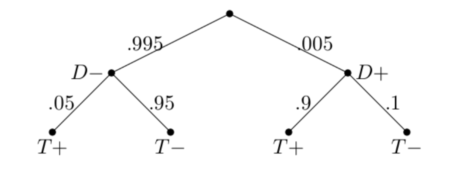
Ejemplo explicativo (1)

$$\mathbb{P}(\text{falso positivo}) = \mathbb{P}(T+|D-) = .95 \quad \text{y} \quad \mathbb{P}(\text{falso negativo}) = \mathbb{P}(T-|D+) = .1.$$

Las probabilidades complementarias se conocen como las tasas negativas y positivas verdaderas:

$$\mathbb{P}(T-|D-) = 1-\mathbb{P}(T+|D-) = 0.95 \quad \text{y} \quad \mathbb{P}(T+|D+) = 1-\mathbb{P}(T-|D+) = .9.$$

Árboles: Todas estas probabilidades se pueden mostrar muy bien en un árbol.



La pregunta pide la probabilidad de que una persona tenga la enfermedad dado que dió positivo, es decir, cuál es el valor de $\mathbb{P}(D+|T+)$.

Ejemplo explicativo (2)

No se nos da este valor, pero si sabemos $\mathbb{P}(T+|D+)$, por lo que podemos usar el teorema de Bayes:

$$\mathbb{P}(D+|T+) = \frac{\mathbb{P}(T+|D+) \cdot \mathbb{P}(D+)}{\mathbb{P}(T+)}.$$

Se dan las dos probabilidades en el numerador. Calculamos el denominador $\mathbb{P}(T+)$ utilizando la ley de probabilidad total. Usando el árbol sólo tenemos que sumar las probabilidades para cada uno de los nodos marcados con $T+$.

$$\mathbb{P}(T+) = .995 \times .05 + .005 \times .9 = .05425$$

Así,

$$\mathbb{P}(D+|T+) = \frac{.9 \times .005}{0.5425} = 0.082949 \approx 8.3\,\%.$$

Ejemplo explicativo (3)

Tablas : Otro truco que es útil para calcular las probabilidades es hacer una tabla. Vamos a repetir el ejemplo anterior usando una tabla construida con 10000 personas totales divididas de acuerdo con las probabilidades en este ejemplo.

Construimos la tabla como sigue. Elegimos un número, digamos 10000 personas y colocamos como el total general en la parte inferior derecha. Usando $\mathbb{P}(D+) = .005$ calculamos que 50 de las 10000 personas están enfermas ($D+$). Del mismo modo 9950 personas son saludables ($D-$). En este punto, la tabla se ve así:

	$D+$	$D-$	total
$T+$			
$T-$			
total	50	9950	10000

Usando $P(T+|D+) = .9$ podemos calcular que el número de personas enfermas que dieron positivo en el 90 % de 50 o 45.

Ejemplo explicativo (4)

Las otras entradas son similares. En este punto la tabla se parece a la tabla de abajo a la izquierda. Finalmente sumamos las filas $T+$ y $T-$ para obtener la tabla completa a la derecha.

	$D+$	$D-$	total
$T+$	45	498	
$T-$	5	9452	
total	50	9950	10000

	$D+$	$D-$	total
$T+$	45	498	543
$T-$	5	9452	9457
total	50	9950	10000

Usando la tabla completa podemos calcular:

$$\mathbb{P}(D+\,|T+) = \frac{|D+\cap T+|}{|T+|} = \frac{45}{543} = 8.3\,\%.$$

Ejemplo explicativo (5)

Símbolos: Para completar, mostramos cómo se ve la solución cuando se escribe directamente en símbolos

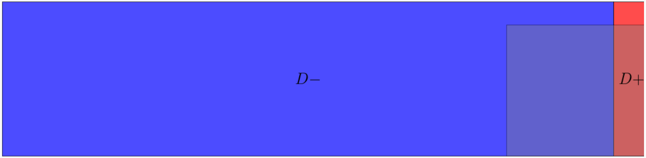
$$
\begin{aligned}
\mathbb{P}(D+|T+) &= \frac{\mathbb{P}(T+|D+)\cdot\mathbb{P}(D+)}{\mathbb{P}(T+)} \\
&= \frac{\mathbb{P}(T+|D+)\cdot\mathbb{P}(D+)}{\mathbb{P}(T+|D+)\mathbb{P}(D+)+\mathbb{P}(T+|D-)\mathbb{P}(D-)} \\
&= \frac{.9\times .005}{.9\times .005+.05\times .995} \\
&= 8.3\,\%
\end{aligned}
$$

Esto se llama la falacia de la tasa de base porque la tasa de base de la enfermedad en la población es tan baja que la gran mayoría de las personas que toman la prueba son saludables, e incluso con una prueba precisa la mayoría de los positivos serán personas sanas.

Visualización

La siguiente figura ilustra la falacia de la tasa de base.

El área azul grande representa a toda la gente sana. El área roja mucho más pequeña representa a los enfermos. El rectángulo sombreado representa a la gente que prueba positivo. El área sombreada cubre la mayor parte de la zona roja y sólo una pequeña parte de la zona azul. Aún así, la mayor parte del área sombreada está sobre azul. Es decir, la mayoria de las pruebas positivas son de personas sanas.



Lectura: Chapter 1 Probability Probability, Markov Chains, Queues and Simulation Willian J.Stewart Princeton University Press 2009.