Análisis de Datos Categóricos

Probabilidad Condicional y Teorema Bayes

Mauricio Bucca
github.com/mebucca
mebucca@uc.cl

15 August, 2024

Probabilidades condicional

Probabilidad condicional

Probabilidad condicional es un concepto crucial en teoría de la probabilidad y subyace al propósito principal del análisis de asociación estadística.

La probabilidad de un evento A después de que nos enteramos de que se ha producido el evento B se denomina probabilidad condicional de A dado B. Formalmente:

$$\mathbb{P}(A \mid B)$$

- Experimento: tirar un dado "justo"
- Espacio muestral, $\Omega : \{1, 2, 3, 4, 5, 6\}$
- A es el evento de obtener un cuatro o más, $A : \{4, 5, 6\}$
- B es el evento de obtener un número par, $B : \{2, 4, 6\}$

Supongamos que tiramos el dado pero no miramos el resultado todavía. Una tercera persona nos dice que obtuvimos un número par.

Pregunta: ¿Cuál es la probabilidad de obtener un cuatro o más una vez que sabemos que el resultado es un número par?

Probabilidad condicional

Formalmente, nuestra pregunta se expresa del siguiente modo: $\mathbb{P}(A \mid B)$.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

Donde,

- $\mathbb{P}(B)$ es la probabilidad de que B ocurra. Es decir obtener un número **par**
- $\mathbb{P}(A, B)$ es la probabilidad de que A y B ocurran conjuntamente. Es decir, obtener un número **par, igual o mayor que 4**

Intuitivamente, queremos saber en qué proporción de los casos en que B ocurre, A también ocurre.

Probabilidad condicional

En nuestro caso,

$$\begin{aligned}
\mathbb{P}(A|B) &= \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(\text{dado=4 o dado=6})}{\mathbb{P}(\text{dado=2 o dado=4 o dado=6})} \\
&= \frac{2/6}{3/6} = \frac{1}{3} \times 2 \approx 0.66
\end{aligned}$$

Probabilidad condicional

Ejercicio rápido:

Supongamos que:

- Un 5% de la población son mujeres (M) con estudios universitarios completos (U)
- Las mujeres representan un 55% de la población
- Un 20% de la población tiene estudios universitarios completos

Pregunta: 1) ¿Cuál es la probabilidad de tener estudios universitarios completos si se es mujer?

Respuesta:

1)

$$\mathbb{P}(U \mid M) = \frac{\mathbb{P}(U, M)}{\mathbb{P}(M)} = \frac{0.05}{0.55} \approx 0.09$$

Probabilidad condicional

Ejercicio rápido:

Supongamos que:

- Un 5% de población son mujeres (M) con un estudios universitarios completos (U)
- Las mujeres representan 55% de la población
- Un 20% de la población tiene estudios universitarios completos

Pregunta: 2) ¿Cuál es la probabilidad de que una persona con estudios universitarios completos sea mujer?

Respuesta:

2)

$$\mathbb{P}(M \mid U) = \frac{\mathbb{P}(U, M)}{\mathbb{P}(U)} = \frac{0.05}{0.2} = 0.25$$

Probabilidad condicional

Ejercicio rápido:

Supongamos que:

- Un 5% de población son mujeres (M) con un estudios universitarios completos (U)
- Las mujeres representan 55% de la población
- Un 20% de la población tiene estudios universitarios completos

Pregunta: 3) ¿Cuál es la probabilidad de que una persona con estudios universitarios completos sea hombre (H)?

Respuesta:

3)

$$\mathbb{P}(H \mid U) = \frac{\mathbb{P}(U, H)}{\mathbb{P}(U)} = 1 - \mathbb{P}(M|U) = 1 - 0.25 = 0.75$$

Probabilidad condicional

Bonus:

4) ¿Cuál es la probabilidad de tener estudios universitarios completos si se es hombre?

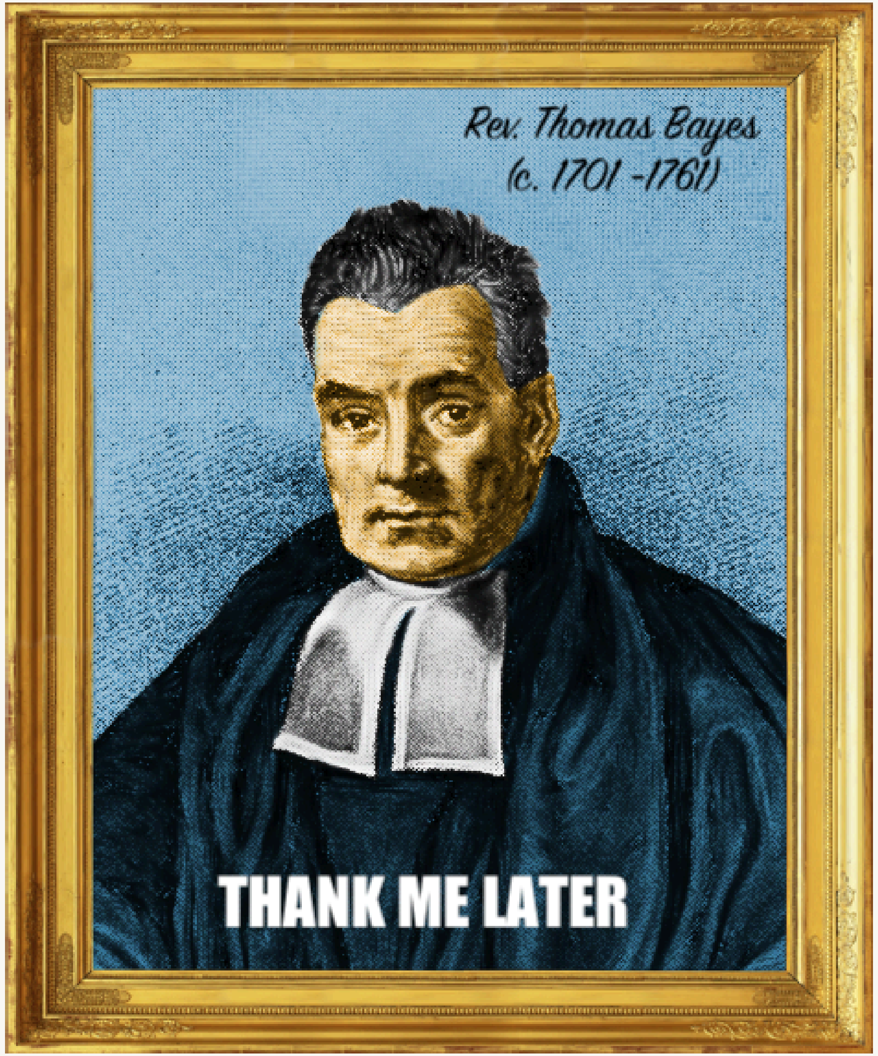
Pista: En general,

$$\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$$

pero el Teorema de Bayes nos dice cómo transformar uno en el otro.

Teorema de Bayes

Thomas Bayes



Teorema de Bayes

La probabilidad de A dado B está definida como:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)} \tag{1}$$

Por tanto, la probabilidad de B dado A está definida como:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(A)} \tag{2}$$

Por tanto:

$$\mathbb{P}(A, B) = \mathbb{P}(B \mid A)\mathbb{P}(A) \tag{3}$$

Reemplazando (3) en (1) obtenemos:

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

Teorema de Bayes

Entonces, si

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

re-ordenando la expresión encontramos ...

Teorema de Bayes:

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


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

Teorema de Bayes en práctica

Ejercicio rápido:

- Un 5% de la población son mujeres (M) con estudios universitarios completos (U)
- Las mujeres representan un 55% de la población
- Un 20% de la población tiene estudios universitarios completos

Bonus: 4) ¿Cuál es la probabilidad de tener estudios universitarios completos si se es hombre?

Respuesta:

4)

$$\mathbb{P}(U \mid H) = \frac{\mathbb{P}(H \mid U)\mathbb{P}(U)}{\mathbb{P}(H)} = \frac{0.75 \times 0.2}{0.45} = 0.333$$

Ley de "probabilidad total"

Ley de "probabilidad total"

Si $\{B_1, B_2., \ldots, B_n\}$ es un conjunto de particiones "desunidas" y mutuamente excluyentes del espacio muestral, entonces:

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

dado que, como vimos, $\mathbb{P}(A, B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$, entonces ...

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

Ley de "probabilidad total"

Ejercicio rápido:

- Las mujeres (M) representan un 55% de la población
- El 20% de los hombres (H) tiene estudios universitarios completos (U)
- El 25% de las mujeres tiene estudios universitarios completos

Pregunta: ¿Cuál es la probabilidad de que una persona seleccionada al azar tenga estudios universitarios completos?

Respuesta:

$$\begin{aligned}\mathbb{P}(U) &= \mathbb{P}(U \mid H)\mathbb{P}(H) + \mathbb{P}(U \mid M)\mathbb{P}(M) \\ &= 0.2 \times 0.45 + 0.25 \times 0.55 \\ &= 0.2275\end{aligned}$$

Independencia estadística

Independencia

Dos eventos A y B son **independientes** cuando la ocurrencia de A no afecta la ocurrencia de B y viceversa.

Independencia es un caso especial de probabilidad condicional: $A \perp B$ si el conocimiento sobre B no cambia nuestro conocimiento sobre A. Formalmente:

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

A partir de esta definición podemos derivar un test matemático de independencia:

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

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

$$\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B)$$

Si dos eventos A y B son independientes entonces debe ser cierto que $\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B)$

Independencia

Ejemplo:

- Experimento: lanzar dos monedas justas consecutivamente
- A es el evento de obtener Cara con la primera moneda
- B es el evento de obtener Sello con la segunda moneda

Pregunta: ¿Son A y B dos eventos independientes?

Sabemos que...

- $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$
- $\Omega : \{CC, CS, SC, SS\}$
- $(A, B) = \{CS\}$
- $\mathbb{P}(A, B) = 1/4$

Test de independencia

Si A y B son eventos independientes entonces deberíamos esperar que $\mathbb{P}(A)\mathbb{P}(B) = 1/4$

$$\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Simulación Monte Carlo

Independencia

Experimento: Tiramos dos monedas simultáneamente, una justa y una moneda cargada (90% prob. Cara)

Si ambos lanzamientos son efectivamente independientes la probabilidad de obtener dos Caras debiera ser: $\mathbb{P}(C_1C_2) = \mathbb{P}(C_1)\mathbb{P}(C_2) = 0.5 \times 0.9 = 0.45$

```r
# lanzamos ambas monedas 100000 veces
coin1 <- rbinom(1000,1, prob = 0.5)
coin2 <- rbinom(1000,1, prob = 0.9)
data_coins <- tibble(coin1=coin1,coin2=coin2)

print(data_coins, n=7)
```

```
## # A tibble: 1,000 × 2
##   coin1 coin2
##   <int> <int>
## 1     0     1
## 2     1     1
## 3     1     1
## 4     1     1
## 5     0     1
## 6     1     1
## 7     1     1
## # ℹ 993 more rows
```

```r
# definimos evento exitoso: dos Caras
data_coins <- data_coins %>%
    mutate(cc = if_else(
      coin1==1 & coin2==1,
      "CC","Otro"))


#  probabilidad de obtener 2 Caras

knitr::kable(prop.table(
  table(data_coins$cc)),
  col.names = c("Resultado","Probabilidad"))
```

Resultado	Probabilidad
CC	0.43
Otro	0.57

Hasta la próxima clase. Gracias!

Mauricio Bucca
https://mebucca.github.io/
github.com/mebucca