Probabilidad

Probabilidad condicional
16/03/2023

Lara Sánchez Peña[1]

UTDT 2023 - MEC

[1]Basado en las notas de Andrea Rotnitzky

Contenido

- Probabilidad condicional
- Ley de probabilidad total
- Regla de multiplicación
- Regla de Bayes
- Eventos independientes

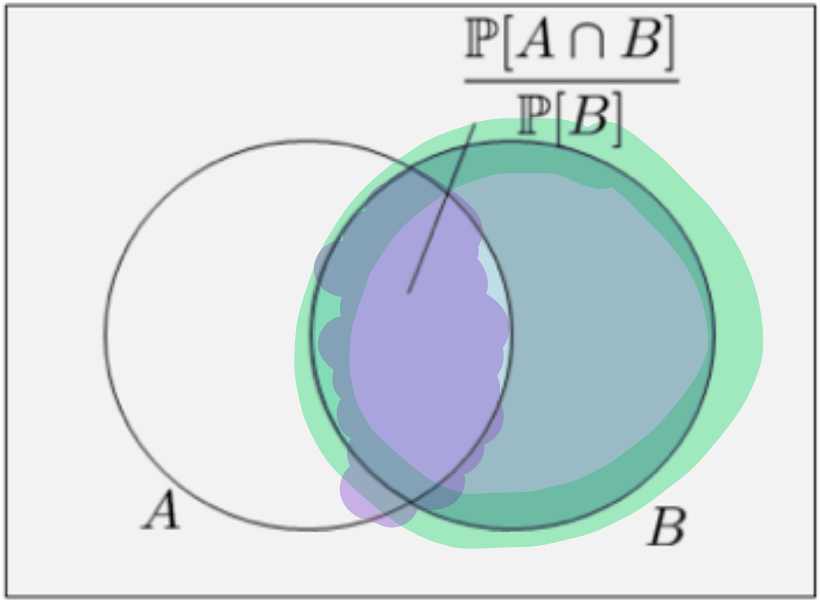
Definición: probabilidad condicional

Sabiendo que ocurre el evento B.

Definimos la **probabilidad condicional** del evento A dado que sabemos que ocurrió el evento B, con $P(B) > 0$ como

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

$$P(B|B) = \frac{P(B \cap B)}{P(B)} = 1$$



Definiciones

espacio muestral = {⚀ ⚁ ⚂ ... ⚅} = Ω

{ϕ, Ω, {⚂}, {⚁ ⚃ ⚄} ...}

Si tenemos un modelo probabilístico $(\Omega, \Sigma, P(\cdot))$ y sabemos que ocurrió el evento A podemos incorporar esa información para obtener un nuevo modelo probabilístico:

supongamos que sabemos que salió al menos ⚁ en el dado

$$(A, \Sigma_A, P(\cdot|A))$$

A = {⚁ ⚂ ⚃ ⚄ ⚅}.

Demuestre que $P(\cdot|A)$ define una probabilidad sobre el nuevo espacio muestral A. Es decir, demuestre que verifica los axiomas de probabilidad.

- $P(A|A) = 1$,
- $P(B_1|A) \geq 0$,
- Para cualesquiera B_1 y B_2 eventos disjuntos que pertenecen a Σ_A vale que $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A)$

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} \in [0,1]$$

Ejemplo 1: Probabilidad condicional

Puntaje crediticio

- Para otorgar un préstamo hipotecario, un banco recolecta información sobre la capacidad de pago de los solicitantes y les asigna un puntaje de 0 a 100.
- El préstamo es otorgado a quienes hayan obtenido un puntaje mayor o igual que 80. No se tiene información adicional sobre cada individuo.
- El banco toma de su base de datos la siguiente información: los puntajes de 100,000 clientes que recibieron el crédito y si resultaron resultaron morosos.

$\text{Prob}(\text{CAT2}) = \frac{22}{100}$

Puntaje (Categorías 1,2,3,4)

		CAT 1	CAT 2	CAT 3	CAT 4	
		80-85	**86-90**	**91-95**	**96-100**	total
Moroso	**no**	8	16	24	32	80
	sí	8	6	4	2	20
	total	16	22	28	34	100

$P(\text{moroso}) = \frac{20}{100}$

$P(\text{cat 2 y moroso}) = \frac{6}{100}$

Cuadro: Tabla de contingencia (en miles de personas).

Ejemplo 1: Probabilidad condicional

Diga cuál es el espacio muestral Ω si elegimos al azar un cliente que recibió un préstamo y responda:

no repaga a tiempo el préstamo.

1. ¿Cuál es la probabilidad de que sea moroso?

$$M = \text{“el cliente elegido es moroso”}$$

$$P(M) = \frac{\#\text{clientes morosos}}{\#\text{ total de clientes}} = \frac{20}{100} = 0{,}2$$

$P(\text{CAT2}) = \frac{22}{100}$

2. ¿Cuál es la probabilidad de que su puntaje esté entre 86 y 90?

$$\text{Cat 2} = \text{“el cliente recibió un puntaje entre 86-90”}$$

$$P(\text{Cat 2}) = \frac{\#\text{ clientes con puntaje entre 86-90}}{\#\text{ total de clientes}} = \frac{22}{100} = 0{,}22$$

3. ¿Cuál es la probabilidad de que el cliente haya recibido un puntaje entre 86 y 90 y además sea moroso?

$$P(M \cap \text{Cat 2}) = \frac{\#\text{ clientes morosos con puntaje entre 86-90}}{\#\text{ total de clientes}} = \frac{6}{100} = 0{,}06$$

Ejemplo 1: Probabilidad condicional

$P(\text{Morosa}) = 0.2$

Supongamos ahora que el banco recibe una solicitud de un **nuevo** cliente interesado en un préstamo hipotecario.

$P(\text{Morosa} \mid \text{Cat } 2) = \frac{6}{22} = 0.\overline{27}$

4. Sabiendo que la solicitud de esa persona recibe un puntaje entre 86-90, ¿cuál es la probabilidad de que, al otorgarle el préstamo, la persona resulte morosa?

$P(\text{Morosa} \mid \text{no cat } 2) = \frac{14}{78} = 0.179487\ldots$

5. Solamente sabiendo que la solicitud recibe un puntaje mayor o igual que 80 pero fuera de la categoría 86-90, ¿cuál es la probabilidad de que, al otorgarle el préstamo, la persona resulte morosa?

Si las últimas dos probabilidades resultaran distintas, **saber si la persona está o no en la categoría** $86-90$ nos permitiría a hacer una mejor predicción sobre la probabilidad de que la persona sea morosa.

Ejemplo 1: Probabilidad condicional

Tabla de contingencia

		Puntaje (Categorías 1,2,3,4)				
		80-85	**86-90**	**91-95**	**96-100**	total
Moroso	**no**	8	16	24	32	80
	sí	8	6	4	2	20
	total	16	22	28	34	100

4. Sabiendo que la solicitud de esa persona recibe un puntaje entre 86-90. ¿Cuál es la probabilidad de que, en caso de otorgársele el prestamo, la persona resulte ser morosa?

$$P(M|\text{Cat } 2) = P\left(\textit{moroso } \textbf{dado que } \textit{tiene puntaje entre 86-90}\right)$$
$$= \frac{\#\text{clientes morosos y con puntaje entre 86-90}}{\#\text{total de clientes con puntaje entre 86-90}} = \frac{6}{22} \approx 0{,}273$$

Ejemplo 1: Probabilidad condicional

Notemos que la probabilidad

$$P(M|\text{Cat } 2) = \frac{\#\text{clientes morosos y con puntaje entre 86-90}}{\#\text{total de clientes con puntaje entre 86-90}}$$

se puede reescribir. Dividiendo tanto numerador como denominador por el número total de clientes,

$$P(M|\text{Cat } 2) = \frac{\frac{\#\text{ clientes morosos y con puntaje entre 86-90}}{\#\text{total de clientes}}}{\frac{\#\text{total de clientes con puntaje entre 86-90}}{\#\text{total de clientes}}}$$

queda que

$$P(M|\text{Cat } 2) = \frac{P(\text{moroso y puntaje entre 86-90})}{P(\text{puntaje entre 86-90})}$$

$$= \frac{P(M \cap \text{Cat } 2)}{P(\text{Cat } 2)} = \frac{6/100}{22/100} = \frac{6}{22}$$

Ejemplo 1: Probabilidad condicional

		Puntaje (Categorías 1,2,3,4)				
		80-85	**86-90**	**91-95**	**96-100**	total
Moroso	**no**	8	16	24	32	80
	sí	8	6	4	2	20
	total	16	22	28	34	100

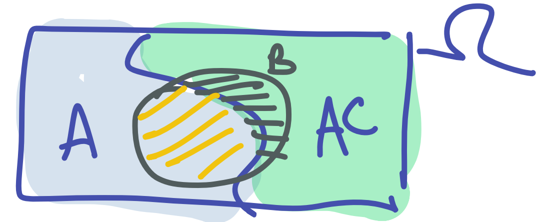
5. Solamente sabiendo que la solicitud recibe un puntaje mayor o igual que 80 pero fuera de la categoría 86-90, ¿cuál es la probabilidad de que, en caso de que se le otorgue el préstamo, la persona resulte morosa?

$$P(M|\text{Cat } 2^c) = \frac{\#\text{clientes morosos con puntaje} \geq 80 \text{ pero no entre 86 y 90}}{\# \text{ total de clientes con puntaje} \geq 80 \text{ pero no entre 86 y 90}}$$
$$= \frac{8+4+2}{16+28+34} = \frac{14}{78} = \frac{7}{39} \approx 0{,}179$$

Notemos entonces que saber si el individuo pertenece a la categoría 2 o no, nos da información sobre si es más probable que el individuo sea moroso.

Definición: Ley de probabilidad total (LPT)

Queremos describir $P(B)$ en como una suma de probabilidades según si ocurre A u ocurre A^c. Observemos que los eventos



$$\underline{\underline{B \cap A}} \quad \text{y} \quad \underline{\underline{B \cap A^c}}$$

son **mutuamente excluyentes** pues si ocurre A no puede ocurrir A^c. En este caso consideramos que Ω se puede particionar como $\Omega = A \mathbin{\dot\cup} A^c$.

Además, podemos escribir al conjunto B como una partición de dos conjuntos disjuntos:

$$(B \cap A) \mathbin{\dot\cup} (B \cap A^c) = B\,.$$

unión disjunta

Entonces,

$$P(B) = P\left[(B \cap A) \mathbin{\dot\cup} (B \cap A^c)\right] = P(B \cap A) + P(B \cap A^c)$$

Con la definición de prob. condicional, obtenemos un caso particular de la LPT:

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Como $\boxed{P(B) = P(B \cap A) + P(B \cap A^c)}$

ó ocurre $A \longrightarrow P(B \cap A) \overset{\circledast_1}{=} P(B|A) \cdot P(A)$

o $\boxed{\text{no ocurre } A} =$ ocurre A^c

$P(B \cap A^c) \overset{\circledast_2}{=} P(B|A^c)\, P(A^c)$

Ley de probabilidad total

$$P(B) = P(B|A)\,P(A) + P(B|A^c)\,P(A^c)$$

$$\frac{P(B \cap A)}{P(A)} = P(B|A) \overset{\circledast_1}{\iff} P(B \cap A) = P(B|A) \cdot P(A)$$

$$\frac{P(B \cap A^c)}{P(A^c)} = P(B|A^c) \underset{\circledast_2}{\iff} P(B \cap A^c) = P(B|A^c)\,P(A^c)$$

Definición: Ley de probabilidad total (LPT)

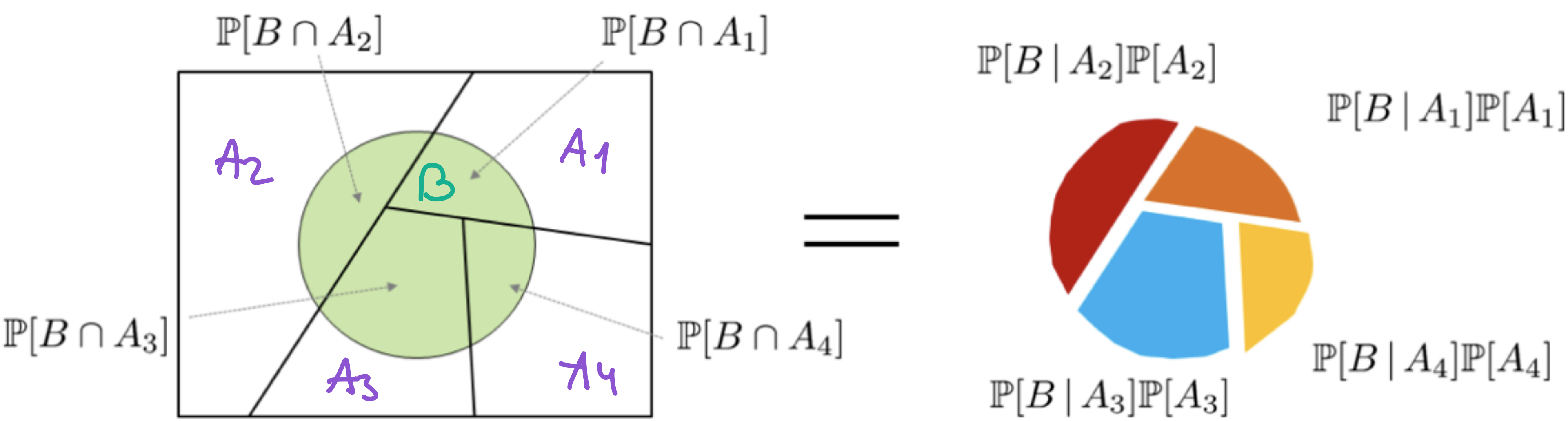
Más en general, si $\{A_1, \ldots, A_n\}$ es una **partición** de Ω, es decir, $\dot{\bigcup}_{i=1}^{n} A_i = \Omega$;

entonces, se tiene que

$\Omega = A_1 \dot{\cup} A_2 \dot{\cup} A_3 \ldots \dot{\cup} A_n$

$$P(B) = \overbrace{P(B|A_1)P(A_1)}^{P(A_1 \cap B)} + \cdots + \overbrace{P(B|A_n)P(A_n)}^{P(A_n \cap B)}$$

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$



$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots +$$
$$+ P(B|A_n) \cdot P(A_n).$$

Ejemplo
$$\begin{cases} B = \text{ser moroso} \\ A_1 = \text{cat } 1 \\ A_2 = \text{cat } 2 \\ A_3 = \text{cat } 3 \\ A_4 = \text{cat } 4 \end{cases}$$

$$\overbrace{P(B)}^{20/100} = P(B|\text{cat}_1) \cdot P(\text{cat}_1)$$
$$+ P(B|\text{cat}_2) \cdot P(\text{cat}_2)$$
$$+ P(B|\text{cat}_3) \cdot P(\text{cat}_3)$$
$$+ P(B|\text{cat}_4) \cdot P(\text{cat}_4)$$

$$= \frac{8}{\cancel{16}} \cdot \frac{\cancel{16}}{100} \quad (= P(B \cap \text{cat } 1))$$
$$+ \frac{6}{\cancel{22}} \cdot \frac{\cancel{22}}{100} \quad (= P(B \cap \text{cat } 2))$$
$$+ \frac{4}{\cancel{28}} \cdot \frac{\cancel{28}}{100}$$
$$+ \frac{2}{\cancel{34}} \cdot \frac{\cancel{34}}{100}$$

Puntaje (Categorías 1,2,3,4)

Moroso		80-85	86-90	91-95	96-100	total
	no	8	16	24	32	80
	sí	8	6	4	2	20
	total	16	22	28	34	100

$$= \frac{8}{100} + \frac{6}{100} + \frac{4}{100} + \frac{2}{100} = \frac{20}{100}.$$

Ejemplo 1: Ley de probabilidad total

En el ejemplo del banco sabemos que $P(M) = 0{,}2$. Supongamos que no lo supiéramos. Si tuviéramos la información de los clientes morosos para cada una de las categorías $P(M|\text{Cat } i)$ podríamos calcular $P(M)$ usando LPT. El conjunto de clientes se puede dividir en cuatro categorías:

Cat 1 (2) 3 (4): si el cliente saca puntaje entre 80-85, (86-90), 91-95, (96-100).

Definiendo

- $\Omega = \{(c_i, s),\ \text{donde } i = 1, 2, 3, 4 \text{ y } s \in \{\text{moroso},\ \text{no moroso}\}\}$,
- $\text{Cat } i = \{(c_i, \text{moroso}); (c_i,\ \text{no moroso})\}$ para $i = 1, 2, 3, 4$ y
- $M = \{(c_1, \text{moroso}); (c_2, \text{moroso}); (c_3, \text{moroso}); (c_4, \text{moroso})\}$

Los eventos Cat 1, Cat 2, Cat 3, Cat 4 constituyen una **partición de Ω** pues

- Son **mutuamente excluyentes**: es decir cualquier par de ellos tiene intersección vacía, pues no pueden ocurrir en simultáneo.
- **La unión de todos ellos constituye todo Ω**: pues un cliente al que se le otorgó un préstamo necesariamente debe caer en alguna de las cuatro categorías.

Ejemplo 1: Ley de probabilidad total

Como $\text{Cat } 1 \uplus \text{Cat } 2 \uplus \text{Cat } 3 \uplus \text{Cat } 4 = \Omega$ entonces,

$$M = (\text{Cat } 1 \cap M) \uplus (\text{Cat } 2 \cap M) \uplus (\text{Cat } 3 \cap M) \uplus (\text{Cat } 4 \cap M)$$

Los eventos

$$(\text{Cat } 1 \cap M), (\text{Cat } 2 \cap M), (\text{Cat } 3 \cap M), (\text{Cat } 4 \cap M)$$

son mutuamente excluyentes, entonces

$$P(M) = P(\text{Cat } 1 \cap M) + P(\text{Cat } 2 \cap M) + P(\text{Cat } 3 \cap M) + P(\text{Cat } 4 \cap M)$$

Por la LPT,

$$\begin{aligned} P(M) =& P(M|\text{Cat } 1)\, P(\text{Cat } 1) + P(M|\text{Cat } 2)\, P(\text{Cat } 2) \\ &+ P(M|\text{Cat } 3)\, P(\text{Cat } 3) + P(M|\text{Cat } 4)\, P(\text{Cat } 4) \\ & \frac{8}{16} \cdot \frac{16}{100} + \frac{6}{22} \cdot \frac{22}{100} + \frac{4}{28} \cdot \frac{28}{100} + \frac{2}{34} \cdot \frac{34}{100} = 0{,}2 \end{aligned}$$

Ejemplo 2: Ley de probabilidad total

Supongamos que entrás a jugar a un torneo de ajedrez, en el cual la probabilidad de que ganes es

- 0,3 cuando te toca jugar contra contrincantes de tipo 1,
- 0,4 cuando te toca jugar contra contrincantes de tipo 2 y
- 0,5 cuando te toca jugar contra contrincantes de tipo 3.

Además, se sabe que 50 % de los contrincantes son de tipo 1, 25 % de los contrincantes son de tipo 2 y el resto son contrincantes de tipo 3.

Se elige **al azar** un contrincante (jugás con cualquier contrincante de manera equiprobable), ¿cuál es la probabilidad de que ganes?

Definimos:

- $\Omega = \{(j_i, r),\ \text{donde } i = 1, 2, 3 \text{ y } r \in \{\text{gané, perdí}\}\}$
- $A_i = \{\text{el contrincante es de tipo } i\} = \{(j_i, \text{perdí}) : (j_i, \text{gané})\};\ i = 1, 2, 3.$
- $B = \{(j_1, \text{gané}); (j_2, \text{gané}); (j_3, \text{gané})\}$

$$P(B) = \overbrace{P(B \mid A_1)}^{0.3}\, \overbrace{P(A_1)}^{0.5} + \overbrace{P(B \mid A_2)}^{0.4}\, \overbrace{P(A_2)}^{0.25}$$

$$+ \underbrace{P(B \mid A_3)}_{0.5}\, \underbrace{P(A_3)}_{0.25} = 0.375$$

Si $P(B \mid A_1) = 0.3$ entonces $P(B^c \mid A_1) = 0.7$

Ejemplo 2: Ley de probabilidad total

Sea A_i, para $i = 1, 2, 3$. Entonces

$$P(A_1) = 0{,}5, \;\; P(A_2) = 0{,}25, \;\; P(A_3) = 0{,}25.$$

Los eventos A_i, $i = 1, 2, 3$ forman una **partición** de Ω.

Sea B el evento de ganar. Entonces

$$P(B|A_1) = 0{,}3, \;\; P(B|A_2) = 0{,}4, \;\; P(B|A_3) = 0{,}5.$$

Luego, aplicando la ley de probabilidad total,

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\ &= 0{,}3 \cdot 0{,}5 + 0{,}4 \cdot 0{,}25 + 0{,}5 \cdot 0{,}25 \\ &= 0{,}375. \end{aligned}$$

La probabilidad de ganar es 0,375.

Regla de multiplicación (dos eventos)

De la definición de probabilidad condicional,

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

podemos despejar $P(A \cap B)$ obteniendo

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Esta fórmula es un caso particular de la **regla de multiplicación**, ya que estamos considerando solamente dos eventos.

En el próximo diagrama de árbol podemos ver la intuición de por qué vale la regla de la multiplicación.

Regla de la multiplicación

Para 3 eventos

$$P(A\cap B\cap C) = P(A\cap(B\cap C))$$

$$= P(A\,|\,B\cap C)\cdot P(B\cap C)$$

$$= P(A\,|\,B\cap C)\cdot \underbrace{P(B|C)\cdot P(C)}_{\circledast_1}$$

$$P(A\cap B\cap C) = P(A\,|\,C\cap B)\cdot P(C\cap B)$$

$$= P(A\,|\,C\cap B)\;\underbrace{P(C|B)\cdot P(B)}_{\circledast_2}$$

Veamos que estas dos expresiones son =.

$\circledast_1$

$$P(B|C)\cdot P(C) = \frac{P(B\cap C)}{\cancel{P(C)}}\cdot \cancel{P(C)} = P(B\cap C)$$

$\circledast_2$

$$P(C|B)\cdot P(B) = \frac{P(C\cap B)}{P(B)}\cdot P(B) = P(C\cap B) = P(B\cap C)$$

Regla de Bayes

$$P(B|C)\cdot P(C) = P(C|B)\cdot P(B)$$

Ejemplo 1: Diagrama de árbol (Ley de probabilidad total)

Visualizamos las probabilidades de eventos que ocurren **secuencialmente**.

$P(\text{Cat } 4) = 0{,}34$

- $P(M^c|\text{Cat } 4) = \frac{32}{34}$ $\quad P(\text{Cat } 4 \cap M^c) = \frac{34}{100} \cdot \frac{32}{34} = 0{,}32$
- $P(M|\text{Cat } 4) = \frac{2}{34}$ $\quad P(\text{Cat } 4 \cap M) = \frac{34}{100} \cdot \frac{2}{34} = 0{,}02$

$P(\text{Cat } 3) = 0{,}28$

- $P(M^c|\text{Cat } 3) = \frac{24}{28}$ $\quad P(\text{Cat } 3 \cap M^c) = \frac{28}{100} \cdot \frac{24}{28} = 0{,}24$
- $P(M|\text{Cat } 3) = \frac{4}{28}$ $\quad P(\text{Cat } 3 \cap M) = \frac{28}{100} \cdot \frac{4}{28} = 0{,}04$

$P(\text{Cat } 2) = 0{,}22$

- $P(M^c|\text{Cat } 2) = \frac{16}{22}$ $\quad P(\text{Cat } 2 \cap M^c) = \frac{22}{100} \cdot \frac{16}{22} = 0{,}16$
- $P(M|\text{Cat } 2) = \frac{6}{22}$ $\quad P(\text{Cat } 2 \cap M) = \frac{22}{100} \cdot \frac{6}{22} = 0{,}06$

$P(\text{Cat } 1) = 0{,}16$

- $P(M^c|\text{Cat } 1) = \frac{8}{16}$ $\quad P(\text{Cat } 1 \cap M^c) = \frac{16}{100} \cdot \frac{8}{16} = 0{,}08$
- $P(M|\text{Cat } 1) = \frac{8}{16}$ $\quad P(\text{Cat } 1 \cap M) = \frac{16}{100} \cdot \frac{8}{16} = 0{,}08$

Regla de multiplicación (tres eventos)

Por ejemplo, supongamos que los clientes son también clasificados de acuerdo a si el crédito otorgado fue por una cifra mayor a 300.000 pesos o no.

Entonces tendremos 3 clasificaciones: a) si el cliente es moroso **(M)** o no, b) si el cliente está en la categoría 2 **(Cat 2)**, c) si el cliente recibió un préstamo de más de 300.000 pesos **(A)** o no.

Notar que en este caso,

$$\Omega = \{(\text{cat } i, s, v) : i = 1, 2, 3, 4 \text{ y } s \in \{\text{m., no m.}\}, v \in \{> \text{ 300k}, \leq \text{ de 300k}\}\}$$

Entonces la probabilidad de que un cliente elegido al azar sea un cliente moroso que recibió un puntaje 86-90 (categoría 2) y un crédito de más de 300.000 pesos es:

$$\begin{aligned} P(A \cap M \cap \text{Cat 2}) &= P(A \cap M | \text{Cat 2}) \cdot P(\text{Cat 2}) \\ &= \underbrace{P(A | M \cap \text{Cat 2})}_{\text{6 personas}} \cdot \underbrace{P(M | \text{Cat 2})}_{6/22} \cdot \underbrace{P(\text{Cat 2})}_{22/100} \end{aligned}$$

¿por qué?

$$P(A \cap M \mid \text{cat } 2) = \frac{P(A \cap M \cap \text{cat } 2)}{P(\text{cat } 2)}$$

$$= \frac{P(A \cap M \cap \text{cat } 2)}{P(\text{cat } 2)} \cdot \frac{P(M \cap \text{cat } 2)}{P(M \cap \text{cat } 2)}$$

$$= P(M \mid \text{cat } 2) \; P(A \mid M \cap \text{cat } 2)$$

Regla de multiplicación (n eventos)

Dados n eventos $A_1, \ldots, A_n$, entonces la regla de multiplicación nos permite hallar la probabilidad de su intersección como:

$$\begin{aligned} P\left(A_1 \cap A_2 \cdots \cap A_n\right) &= P\left(A_n \middle| A_1 \cap A_2 \cdots \cap A_{n-1}\right) \cdot \\ &\quad P\left(A_{n-1} \middle| A_1 \cap A_2 \cdots \cap A_{n-2}\right) \cdots \\ &\quad P\left(A_2 \middle| A_1\right) \cdot P\left(A_1\right) \end{aligned}$$

productoria ("multiplicatoria")

De manera sucinta,

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k \middle| \bigcap_{j=1}^{k-1} A_j\right)$$

La regla es útil en muchas situaciones donde el cálculo de cada probabilidad condicional es relativamente fácil e intuitivo mientras que el cálculo directo de $P\left(A_1 \cap A_2 \cdots \cap A_k\right)$ puede no ser tan sencillo. **Nota**: $A_0 = \Omega$.

Regla de multiplicación: un ejemplo

Una clase tiene 4 estudiantes de posgrado y 12 de grado. Se los divide al azar en cuatro grupos de cuatro integrantes, ¿cuál es la probabilidad de que cada grupo tenga exactamente un estudiante de posgrado? **(Note que el orden de los grupos no es relevante**.

Enumeremos los estudiantes de posgrado como P, Q, R, y S. Consideremos los siguientes eventos

- $A_1 = \{\text{Los estudiantes P y Q están en grupos distintos}\}$
- $A_2 = \{\text{Cada uno de los estudiantes P, Q y R está en un grupo distinto}\}$
- $A_3 = \{\text{Cada uno de los estudiantes P, Q, R y S está en un grupo distinto}\}$

Entonces

$$A_3 = A_3 \cap A_2 \cap A_1$$

Por lo tanto, por la regla de multiplicación para tres eventos:

$$P(A_3) = P(A_3 \cap A_2 \cap A_1) = \overbrace{P(A_3 | \underbrace{A_2 \cap A_1}_{A_2})}^{4/13} \cdot \overbrace{P(A_2 | A_1)}^{8/14} \cdot \overbrace{P(A_1)}^{12/15}$$

Regla de multiplicación: un ejemplo

- $P(A_1) = \dfrac{12}{15} = \dfrac{4}{5}$ porque

$$\underbrace{\times\ \times\ \times P}_{\text{Grupo 1}} \quad \underbrace{-\ -\ -\ -}_{\text{Grupo 2}} \quad \underbrace{-\ -\ -\ -}_{\text{Grupo 3}} \quad \underbrace{-\ -\ -\ -}_{\text{Grupo 4}}$$

- $P(A_2 | A_1) = \dfrac{8}{14} = \dfrac{4}{7}$

$$\underbrace{\times\ \times\ \times P}_{\text{Grupo 1}} \quad \underbrace{-\ -\ -\ -}_{\text{Grupo 2}} \quad \underbrace{-\ -\ -\ -}_{\text{Grupo 3}} \quad \underbrace{\times\ \times\ \times Q}_{\text{Grupo 4}}$$

- $P(A_3 | A_2 \cap A_1) = P(A_3 | A_2) = \dfrac{4}{13}$

$$\underbrace{\times\ \times\ \times P}_{\text{Grupo 1}} \quad \underbrace{-\ -\ -\ -}_{\text{Grupo 2}} \quad \underbrace{R \times\ \times\ \times}_{\text{Grupo 3}} \quad \underbrace{\times\ \times\ \times Q}_{\text{Grupo 4}}$$

Luego,

$$P(A_3) = \frac{12}{15} \cdot \frac{8}{14} \cdot \frac{4}{13} \approx 0{,}14066$$

Definición: Regla de Bayes

La regla de Bayes relaciona las probabilidades de los eventos $P(A) \neq 0$, $P(B) \neq 0$, $P(A|B)$ y $P(B|A)$.

$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

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

Para demostrar este resultado usamos la definición de probabilidad condicional, empezando por la siguiente igualdad:

$$P(A \cap B) = P(A \cap B)$$

Multiplicamos por $\dfrac{1}{P(B)}$ a ambos lados y del lado derecho multiplicamos y dividimos por $P(A)$:

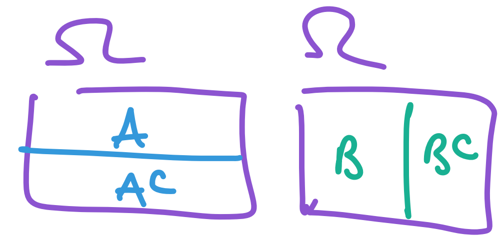
$$\frac{P(A \cap B)}{P(B)} = P(A) \cdot \frac{P(A \cap B)}{P(A)} \frac{1}{P(B)}$$

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

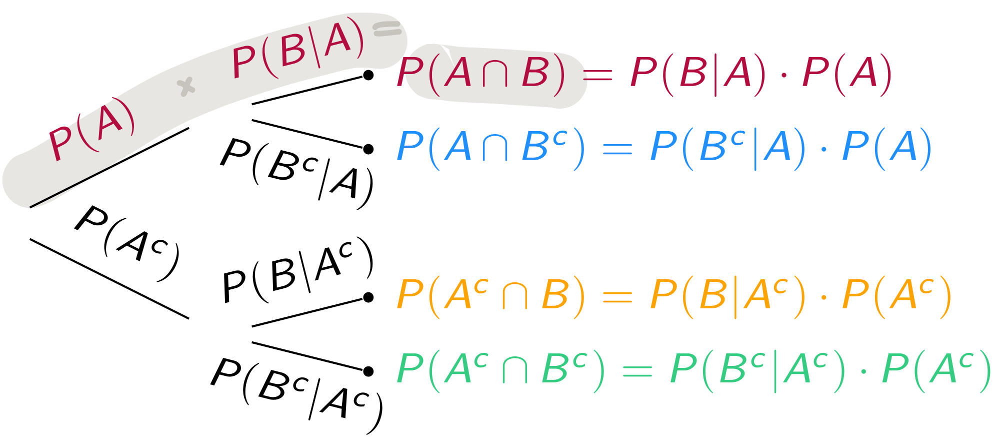
Demostración: Regla de Bayes

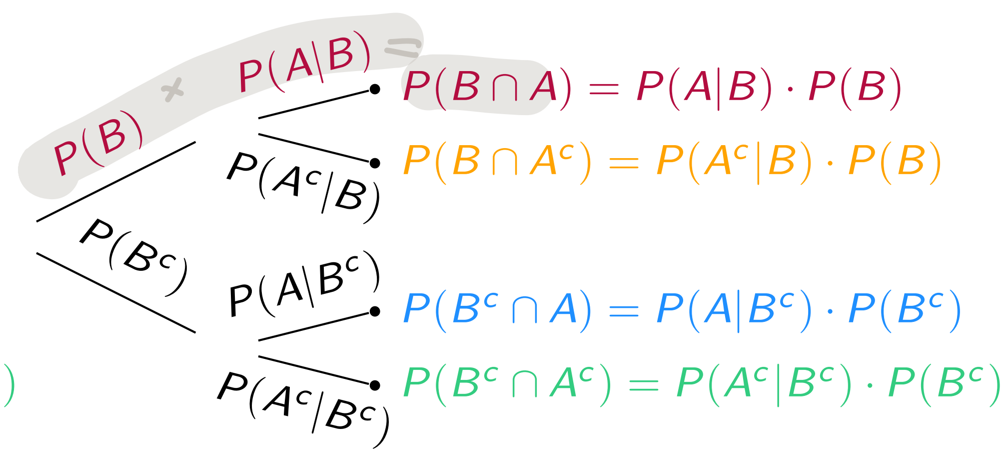
En la inferencia frecuentista la regla de Bayes nos dice que no debería importar el orden en el que condicionamos si ocurren los eventos A y B o no. Reescribiendo la regla de Bayes tenemos que:

$$P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$$



Veámoslo usando dos diagramas de árbol. En el de la izquierda condicionamos primero a si el evento A ocurre o si no ocurre; mientras que, en el de la derecha, condicionamos primero a si el evento B ocurre o si no ocurre.





Regla de Bayes caso general

$A_1 \dot\cup A_2 \dot\cup A_3 \dot\cup \cdots \dot\cup A_n = \Omega$

Sean eventos A_i de manera que $\dot{\bigcup}_{i=1}^{n} A_i = \Omega$ (una partición de Ω) y B

Entonces, usando la regla de Bayes para dos eventos, sabemos que

$$P(A_i|B) = P(A_i) \cdot \frac{P(B|A_i)}{P(B)} \quad \text{(Bayes)}$$

Por LPT también sabemos que

$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) P(A_2) + \cdots + P(B|A_n) P(A_n)$ (LPT)

$$P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i)$$

Lo reemplazamos en la primera ecuación y queda que:

$$P(A_i|B) = P(A_i) \cdot \frac{P(B|A_i)}{\sum_{i=1}^{n} P(B|A_i) P(A_i)}$$

Regla de Bayes: Ejemplo 1

Retomemos el ejemplo de la cartera de clientes del banco:

- Si un nuevo solicitante tuvo un puntaje mayor o igual a 80 entonces la probabilidad de que sea moroso es $P(M) = 0{,}2$.
- Si nos enteramos que su puntaje estuvo en la categoría $86 - 90$, entonces la probabilidad de que sea moroso es $P(M|\text{Cat } 2) = \frac{6}{22}$

Al enterarnos que ocurrió el evento Cat 2, **actualizamos** las probabilidades

$$P\left(M\right) \longrightarrow P\left(M|\text{Cat } 2\right)$$

- $P\left(M\right)$ es la probabilidad **a priori**, que cuantifica la probabilidad de que un cliente elegido al azar resultará moroso antes de saber algo sobre su puntaje.
- $P\left(M|\text{Cat } 2\right)$ es la probabilidad **a posteriori**, que cuantifica la probabilidad de que un cliente elegido al azar resulte moroso luego de haber recibido la información de que el evento Cat 2 ocurrió, es decir, luego que se clasificara al cliente con un puntaje $86 - 90$.

Regla de Bayes: Ejemplo 1

La **regla de Bayes** relaciona la **probabilidad a priori** $P(M)$ con la **probabilidad a posteriori** $P(M|\text{Cat }2)$ y nos dice el cálculo para actualizarla usando el **cociente de verosimilitud** $\frac{P(\text{Cat }2|M)}{P(\text{Cat }2)}$:

$$P(M|\text{Cat }2) = P(M) \cdot \frac{P(\text{Cat }2\ |M)}{P(\text{Cat }2)}$$

¿cómo lo demostramos?

- Por la definición de probabilidad condicional:

$$P(M|\text{Cat }2) = \frac{P(\text{Cat }2 \cap M)}{P(\text{Cat }2)}$$

- Aplicamos la regla de multiplicación al numerador y reacomodando términos:

$$P(M|\text{Cat }2) = \frac{P(\text{Cat }2|M) \cdot P(M)}{P(\text{Cat }2)} = P(M) \cdot \frac{P(\text{Cat }2|M)}{P(\text{Cat }2)}$$

Regla de Bayes: Ejemplo 1

Actualizamos la probabilidad de que un cliente sea moroso usando Bayes al conocer en qué categoría cae su puntaje. Tabla de probabilidades a priori:

		Puntaje				
		80-85	**86-90**	**91-95**	**96-100**	total
Moroso	**no**	8	16	24	32	80
	sí	8	6	4	2	20
	total	16	22	28	34	100

Tabla de probabilidades a posteriori recordando que $P(M) = 0{,}2$:

	Puntaje			
	80-85	**86-90**	**91-95**	**96-100**
$P(\text{Cat } i)$	$16/100$	$22/100$	$28/100$	$34/100$
$P(\text{Cat } i \vert M)$	$8/20$	$6/20$	$4/20$	$2/20$
$\frac{P(\text{Cat } i)}{P(\text{Cat } i \vert M)}$	2,5	1,3636	0,7143	0,294
$\frac{P(M)}{P(\text{Cat } i)} \cdot P(\text{Cat } i \vert M)$	0,5	0,273	0,142	0,059
$= P(M \vert \text{Cat } i)$	$= 0{,}2 \cdot 2{,}5$	$= 0{,}2 \cdot 1{,}3636$	$= 0{,}2 \cdot 0{,}7143$	$= 0{,}2 \cdot 0{,}294$

Regla de Bayes: Ejemplo 1

	Puntaje			
	80-85	**86-90**	**91-95**	**96-100**
$P(\text{Cat } i)$	$16/100$	$22/100$	$28/100$	$34/100$
$P(\text{Cat } i \vert M)$	$8/20$	$6/20$	$4/20$	$2/20$
$\frac{P(\text{Cat } i)}{P(\text{Cat } i \vert M)}$	2,5	1,3636	0,7143	0,294
$\frac{P(M)}{P(\text{Cat } i)} \cdot P(\text{Cat } i \vert M)$	0,5	0,273	0,142	0,059
$= P(M \vert \text{Cat } i)$	$= 0{,}2 \cdot 2{,}5$	$= 0{,}2 \cdot 1{,}3636$	$= 0{,}2 \cdot 0{,}7143$	$= 0{,}2 \cdot 0{,}294$

- La actualización de las probabilidades de morosidad:
  - Si nos enteramos que el puntaje está en la categoría $96 - 100$, la probabilidad **a priori** de 0,2 se actualiza **a posteriori** a 0,059.
  - Si nos enteramos que el puntaje está en la categoría $80 - 85$, la probabilidad **a priori** de 0,2 se actualiza **a posteriori** a 0,5.
- La probabilidad de que una persona morosa tenga un puntaje alto disminuye cuanto más alto es el puntaje.

Regla de Bayes: Ejemplo screening test

Consideremos una enfermedad rara, en donde la probabilidad de que una persona elegida al azar tenga la enfermedad es de 0.001 (1 de cada 1000 personas tienen la enfermedad).

$P(A) = 0.001$

Supongamos que un *screening test* que se recomienda a todas las personas:

- tiene una **sensibilidad** del 95 %: si la persona tiene la enfermedad, el 95 % de las veces el análisis da positivo. Es decir, se detecta la enfermedad en un 95 % de los casos. $P(B|A) = 0{,}95$. $\Rightarrow P(B^c|A) = 0.05$
- tiene una **especificidad** del 99 %: si la persona **no** tiene la enfermedad, el 99 % de las veces el análisis da negativo. Es decir, se detecta la ausencia de enfermedad en un 99 % de los casos. $P(B^c|A^c) = 0{,}99$. $P(B|A^c) = 0.01$

Si una persona se realiza el análisis y le da positivo, ¿cuál es la probabilidad de que tenga la enfermedad?

$P(A|B) = ?$

ENFERMEDAD "RARA" $P(A) = 0.001$

Supongamos que tenemos

100 000 personas

100 enfermos (A) → a 95 de ellas les da $\oplus$ el test.

99900 sanos (A^c) → $0.01 \cdot 99900 =$ 999 de ellas les da $\oplus$ el test.

estar enfermo

el test da $\oplus$

$$P(A \mid B) = \frac{95}{95 + 999} \approx 0.1$$

"ENFERMEDAD COMÚN" $P(A) = 0.5$

100000

50000 enfermos → al 95% de las 50000 les da $\oplus$ $\boxed{47500}$

50000 sanos → al 1% de las 50000 les da $\oplus$ $\boxed{500}$

$$P(A \mid B) = \frac{47500}{47500 + 500} \approx 0.989\ldots$$

Más formalmente

$P(A) = 0.001$

$P(B \mid A) = 0.95$ (SENSIBILIDAD)

$P(B^c \mid A^c) = 0.99$ (ESPECIFICIDAD)

Sufras de la enfermedad

si el test dio positivo

¿ $P(A \mid B)$?

Por Bayes

$$P(A \mid B) \cdot \underbrace{P(B)}_{0.95 \cdot 0.001 + 0.01 \cdot 0.999} = \underbrace{P(B \mid A)}_{0.95} \cdot \underbrace{P(A)}_{0.001}$$

por LPT

$$P(B) = \underbrace{P(B \mid A)}_{0.95} \cdot \underbrace{P(A)}_{0.001} + \underbrace{P(B \mid A^c)}_{1-0.99} \cdot \underbrace{P(A^c)}_{= 1 - P(A) = 1 - 0.001}$$

$$P(A \mid B) = \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.01 \cdot 0.999} \approx 0.1$$

Regla de Bayes: screening test

Definimos $\Omega = \{(\text{e, p}); (\text{e, n}); (\text{s, p}); (\text{s, n})\}$, donde "e" y "s" representan tener la enfermedad o estar sano, mientras que "p" y "n" significan que el test da positivo o negativo.
Consideramos los eventos:

(a) $A = \{(\text{e, p}); (\text{e, n})\}$, el evento de estar enfermo.

(b) $B = \{(\text{e, p}); (\text{s, p})\}$, el evento de dar positivo.

- Antes de ver el resultado del test, la probabilidad **a priori** de que la persona tenga la enfermedad es
$$P(A) = 0{,}001$$
- Una vez conocido el resultado del examen, actualizamos la probabilidad de que la persona esté enferma usando la regla de Bayes:
$$P(A|B) = P(A) \cdot \frac{P(B|A)}{P(B)} = 0{,}001 \cdot \frac{0{,}95}{P(B)}$$

✓

Regla de Bayes: screening test

- Para calcular $P(B)$ usamos la ley de probabilidad total. Usando que

$$A^c = \{(\mathsf{s},\ \mathsf{p}); (\mathsf{s},\ \mathsf{n})\}$$

entonces

$$\begin{aligned} P(B) &= P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= 0{,}95 \cdot 0{,}001 + (1 - 0{,}99) \cdot (1 - 0{,}001) \\ &= 0{,}01094 \end{aligned}$$

- Luego,

$$P(A|B) = P(A) \cdot \frac{P(B|A)}{P(B)} = 0{,}001 \cdot \frac{0{,}95}{0{,}01094} = 0{,}001 \cdot 86{,}837 \approx 0{,}09$$

Regla de Bayes: screening test (intuición)

- Como la probabilidad de tener la enfermedad $P(A) = 0{,}001$, una persona de las mil tiene la enfermedad y con probabilidad 0,95 el test da positivo.
- Ahora bien, con probabilidad 99,9 % una persona está sana y hay "9.99" (1 % de 999) personas sanas cuyo test da positivo.
- De esa manera, si tomamos una muestra de 100000 personas, esperamos que de las personas que tienen la enfermedad y de las personas están sanas haya 95 y 999 que tengan resultado positivo respectivamente. Entonces,

$$P(A|B) = \frac{95}{999 + 95} \approx 0{,}09$$

- Esta probabilidad es baja porque la mayoría de las personas que se testean son sanas y la especificidad no es suficientemente alta para descartar que si alguien está sano, el test de negativo.
- Rehaciendo el ejercicio con $P(A) = 0{,}5$ por ejemplo para una enfermedad pandémica (use su imaginación) verá que $P(A|B) = 0{,}989$. Es decir, testear para este tipo de enfermedad es razonable asumir que si una testea positivo probablemente esté enferma.

Regla de Bayes: odds a priori y posteriori

ej: si $P(A) = \frac{1}{10}$ odds de $A = \frac{\frac{1}{10}}{\frac{9}{10}} = \frac{1}{9}$.

A: gana el caballo Belinda

Definición

Al cociente entre la probabilidad de un evento y la probabilidad del complemento del evento, se lo denomina **odds**.

$$\text{odds a priori para el evento } A = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}$$

$$\text{odds a posteriori para el evento } A \text{ dado que ocurre } B = \frac{P(A|B)}{P(A^c|B)}$$

Ejemplo del banco

$$\text{odds (a priori) de que el cliente sea moroso} = \frac{P(M)}{P(M^c)} = \frac{0{,}2}{0{,}8} = \frac{1}{4}$$

Esto nos dice que si un cliente clasificado para un préstamo, a priori, será cuatro veces menos probable que resulte moroso a que sea buen pagador.

Regla de Bayes: odds a priori y posteriori

Si nos enteramos que el puntaje del cliente está en la categoría 2, con puntaje 86-90, el odds a posteriori dado que ocurrió el evento Cat 2:

$$\text{odds (a posteriori) de morosidad} = \frac{P(M|\text{Cat } 2)}{P(M^c|\text{Cat } 2)}$$
$$= \frac{0{,}273}{1-0{,}273} = \frac{1}{727/273} \approx \frac{1}{2{,}66} \approx 0{,}3755$$

De modo que conociendo que el cliente pertenece a la categoría 2 concluimos que el 37.55 % de los casos serán morosos. Decimos que es 2.66 veces más probable que un cliente de la cateogoría 2 no sea moroso a que lo sea.

En el odds a priori si **no conocemos en qué categoría está el puntaje del cliente**, decimos que es 4 veces menos probable que el cliente sea moroso a que sea buen pagador. Ahora bien, **si sabemos que el cliente tiene un puntaje que cae en la categoría 2**, sabemos que es solamente 2.66 veces menos probable que el cliente sea moroso a que sea buen pagador.

Definición: Eventos independientes

Dos eventos A y B, si $0 < P(B) < 1$, son **independientes** cuando

$$P(A|B^c) = P(A|B)$$

Las siguientes afirmaciones son equivalentes:

1. Los eventos A y B son independientes, es decir, $P(A|B) = P(A|B^c)$.
2. $P(A) = P(A|B)$
3. $P(A \cap B) = P(A) \cdot P(B)$

1. $\Rightarrow$ 2. Sabemos por LPT y 1. sabemos que

$$P(A) \underbrace{=}_{LPT} P(A|B)P(B) + P(A|B^c)P(B^c)$$

$$P(A) = P(A|B)P(B) + P(A|B)P(B^c)$$

$$P(A) = P(A|B)\underbrace{[P(B) + P(B^c)]}_{=1}$$

$$P(A) = P(A|B)$$

Demostraciones: Eventos independientes

2. $\Rightarrow$ 1.
Sabemos por LPT y 2. sabemos que

$$
\begin{aligned}
P(A|B) \overbrace{=}^{2.} P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\
P(A|B) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\
P(A|B) - P(A|B)P(B) &= P(A|B^c)P(B^c) \\
P(A|B)\underbrace{(1 - P(B))}_{P(B^c)} &= P(A|B^c)P(B^c) \\
P(A|B) &= P(A|B^c)
\end{aligned}
$$

Notar que podemos cancelar $P(B^c)$ porque como $P(B) < 1$, vale que $P(B^c) > 0$.

Demostraciones: Eventos independientes

2. $\Rightarrow$ 3.: De la definición de probabilidad condicional, tenemos:

$$
\begin{aligned}
P(A) \overbrace{=}^{2.} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A) &= \frac{P(A \cap B)}{P(B)} \\
P(A)P(B) &= P(A \cap B)
\end{aligned}
$$

3. $\Rightarrow$ 2.: Usamos la definición de probabilidad condicional en el paso $(*)$.

$$
\begin{aligned}
P(A)P(B) &= P(A \cap B) \\
P(A) &= \frac{P(A \cap B)}{P(B)} \\
P(A) &= \frac{P(A \cap B)}{P(B)} \underbrace{=}_{(*)} P(A|B)
\end{aligned}
$$

Propiedades: Eventos independientes

Valen las siguientes propiedades (pruébelas):

1. Si A y B son independientes, entonces A y B^c; A^c y B^c: A^c y B son eventos independientes. Lo demostramos para el caso A y B^c, sabiendo que A y B son independientes. Notemos que

   $$P(A) = P(A \cap B) + P(A \cap B^c) = P(A)P(B) + P(A \cap B^c)$$

   Entonces,

   $$\begin{aligned} P(A \cap B^c) &= P(A) - P(A)P(B) \\ P(A \cap B^c) &= P(A)[1 - P(B)] \\ P(A \cap B^c) &= P(A)P(B^c) \end{aligned}$$

2. El evento $\emptyset$ es independiente de cualquier evento A.

   $$0 = P(\emptyset) = P(A \cap \emptyset) = P(\emptyset) \cdot P(A) = 0 \text{ para cualquier } A$$

3. Si $A \subseteq B$, no pueden ser independientes, excepto que $P(B) = 1$.

   $A \cap B = A$

   $$P(A \cap B) = P(A) \neq P(A) \cdot P(B) \text{ si } P(B) \neq 1$$

Si $A \cup B$ (si A y B son disjuntos
o también llamados eventos mutuamente
nos dice excluyentes).

$A \cap B = \phi$

$P(A \cap B) = P(\phi) = 0$

$P(A) \cdot P(B)$ es igual a cero solamente si $P(A) = 0$ o $P(B) = 0$

Si $P(A) \neq 0$ y $P(B) \neq 0$ $P(A) \cdot P(B) \neq 0$.

O sea que si $A \cap B \neq 0$ y $P(A) \neq 0$ y $P(B) \neq 0$ $\Rightarrow$ A y B no pueden ser independientes

Porque $\underbrace{P(A) P(B)}_{\neq 0}$ y $P(A \cap B) = 0$.

Eventos dependientes: Ejemplo 1

Clients de 4 CATEGORÍAS.

Recordemos que la probabilidad de ser moroso es:

$$P(M) = 0{,}2$$

Si sabemos que el cliente no está en Cat 2, entonces la probabilidad de ser moroso es:

$$P(M|\text{Cat } 2^c) = \frac{14}{78} = \frac{7}{39}$$

Si el cliente se encuentra en la Cat 2, entonces la probabilidad de ser moroso es:

$$P(M|\text{Cat } 2) = \frac{6}{22} = \frac{3}{11}$$

Entonces, el evento M es un **evento dependiente** del evento Cat 2 pues conocer si el cliente pertenece o no a dicha categoría nos da información sobre la probabilidad de que éste sea moroso o no.

$$0{,}1795 = P(M|\text{Cat } 2^c) \neq P(M|\text{Cat } 2) = 0{,}2727\ldots$$

Eventos independientes: Ejemplo 2

Supongamos que la tabla de contingencia hubiese sido la siguiente:

		Puntaje				
			CAT2			
		80-85	**86-90**	**91-95**	**96-100**	total
Moroso	**no**	32	24	16	8	80
	sí	8	6	4	2	20
	total	40	30	20	10	100

$$P(M|\textbf{Cat 2}) = \frac{\#\text{ morosos con puntaje entre 86 y 90}}{\#\text{ clientes con puntaje entre 86 y 90}} = \frac{6}{30} = 0{,}2$$

$$P(M|\textbf{Cat 2}^c) = \frac{\#\text{morosos con puntaje}\geq 80\text{ pero no entre 86 y 90}}{\#\text{clientes con puntaje}\geq 80\text{ pero no entre 86 y 90}} = \frac{8+4+2}{40+20+10} = 0{,}2$$

Saber si el cliente pertenece a Cat 2 o no, no nos cambia la probabilidad de que el cliente sea moroso. M. Decimos que son **eventos independientes**:

$$P(M|\text{Cat } 2) = P(M|\text{Cat } 2^c)$$

Eventos independientes: Ejemplo 3

De la población de habitantes mayores de edad de C.A.B.A. elijo al azar una persona. Consideremos los eventos:

- $A = \{\text{personas cuyo último dígito del DNI es par}\}$
- $B = \{\text{personas que están desempleadas}\}$
- $P(B|A)$ es la probabilidad de que una persona elegida al azar, entre las personas con último dígito del DNI par, esté desempleada.
- $P(B)$ es la probabilidad de que una persona elegida al azar esté desempleada.

¿tenemos razones para suponer que $P(B|A) \neq P(B)$?
No hay razón para suponer que esa proporción sea distinta a la proporción de personas desempleadas entre todos los habitantes mayores de edad de C.A.B.A.

Luego, parece razonable suponer que A y B son eventos independientes. Dicho de otro modo, conocer el último digito del DNI de una persona no me ayuda a mejorar mi predicción sobre su estado de desempleo o no desempleo.

Eventos *independientes* contra *mutuamente excluyentes* eventos

Se dice que dos eventos A y B son **mutuamente excluyentes**. Eso quiere decir que $A \cap B = \emptyset$ y, por lo tanto, $P(A \cap B) = 0$. A

Sean dos eventos A y B **mutuamente excluyentes** y que cumplen que $P(A) > 0$ y $P(B) > 0$. ¿pueden ser A y B eventos independientes?

Si A y B fueran independientes tendría que valer que

$$P(A \cap B) = P(A)P(B)$$ B

Sin embargo, sabemos que el lado izquierdo es igual a cero, mientras que el lado derecho es positivo. Entonces,
si se consideran eventos A y B con probabilidad positiva y mutuamente excluyentes, no pueden ser independientes.

Dicho de otro modo, la información que nos provee conocer que ocurrió el evento B nos ayuda a predecir la ocurrencia de un evento A, que es mutuamente excluyente con B, ya que sabemos con certeza que es imposible que ocurra A.

Si ocurren A y B

$P(A) \cdot P(B) = 0.$

$\ominus$ no excluyente

necesariamente $P(A) = 0$ $\ominus$ $P(B) = 0$

$[\ominus \; P(A) = 0$ y $P(B) = 0]$

Independencia de más de dos eventos

Imagine que usted es el ministro de salud pública y debe implementar un protocolo con los pasos que deben seguir los médicos para diagnosticar alergias.

Los análisis clínicos para diagnosticar alergias son costosos, entonces se trata de elegir una estrategia inteligente que permita minimizar costos y posibilidades de error en el diagnóstico.

Nos preguntamos si merece la pena hacer ciertos análisis clínicos más baratos de otras dolencias que nos ayuden a afinar el pronóstico de quién puede sufrir alergias.

Independencia de dos o más eventos

A: tener alergia C: tener colesterol.
B: tener diabetes

Tabla de contingencia

		Colesterol		
		sí	**no**	total
Diabetes	**sí**	8(6)	12(2)	20(8)
	no	32(10)	48(22)	80(32)
	total	40(16)	60(24)	100(40)

Cuadro: La tabla está expresada en unidades de 100000. Los números entre paréntesis son las cifras de individuos que sufren alergia correspondientes a cada celda.

Independencia de dos o más eventos

Consideremos, para un adulto elegido al azar de la población de Argentina, los siguientes tres eventos:

A: la persona sufre alergias

B: la persona es diabética

C: la persona tiene colesterol alto

Algunas preguntas:

- Son $P(A)$ y $P(A|B)$ iguales?
- Son $P(A)$ y $P(A|C)$ iguales?
- Son $P(A)$ y $P(A|B \cap C)$ iguales?
- Son $P(A)$ y $P(A|B^c \cap C)$ iguales?
- Son $P(A)$ y $P(A|B \cap C^c)$ iguales?
- Son $P(A)$ y $P(A|B^c \cap C^c)$ iguales?

Independencia de dos o más eventos

- $P(A) = \dfrac{\#\text{ adultos con alergia}}{\#\text{ adultos}} = \dfrac{40}{100} = 0{,}4$
- $P(A|B) = \dfrac{\#\text{ adultos con alergia y diabetes}}{\#\text{ adultos con diabetes}} = \dfrac{8}{20} = 0{,}4$
- $P(A|C) = \dfrac{\#\text{ adultos con alergia y colesterol alto}}{\#\text{ adultos con colesterol alto}} = \dfrac{16}{40} = 0{,}4$
- $P(A|B \cap C) = \dfrac{\#\text{ adultos con alergia, diabetes y colesterol alto}}{\#\text{ adultos con diabetes y colesterol alto}} = \dfrac{6}{8} = 0{,}75$
- $P(A|B^c \cap C) = \dfrac{\#\text{ adultos con alergia, sin diabetes y con colesterol alto}}{\#\text{ adultos sin diabetes y con colesterol alto}} = \dfrac{10}{32} = 0{,}31$
- $P(A|B \cap C^c) = \dfrac{\#\text{ adultos con alergia, con diabetes y sin colesterol alto}}{\#\text{ adultos con diabetes y sin colesterol alto}} = \dfrac{2}{12} = 0{,}17$

  diabetes no colesterol
- $P(A|B^c \cap C^c) = \dfrac{\#\text{adultos con alergia, sin diabetes y sin colesterol alto}}{\#\text{adultos sin diabetes y sin colesterol alto}} = \dfrac{22}{48} = 0{,}46$
- En este ejemplo estamos en una situación en la que los eventos A, B y C **no son <u>simultáneamente</u> independientes**, a pesar de que los pares de eventos (A, B) y (A, C) sí son independientes de a pares.

Independencia de dos o más eventos

$A_1, A_2, \ldots, A_n$ son independientes

Definición

Si se verifica que para cada subconjunto $S \subseteq \{1, 2, \ldots, n\}$ vale

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$$

$$P(A_i \cap A_j \cap \ldots) = P(A_i) \cdot P(A_j) \cdots$$

se dice que los eventos $A_1, \ldots, A_n$ son independientes (donde $i, j \in S$).

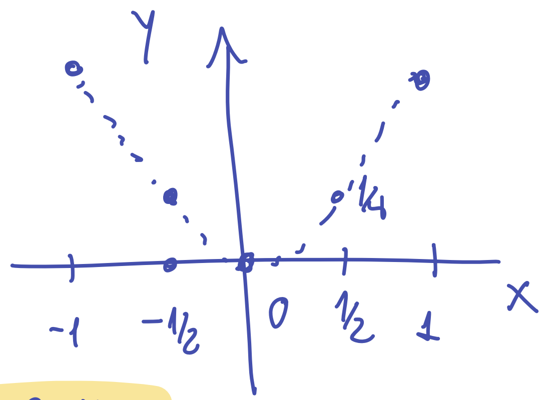
Nota: Es importante recordar que para demostrar que vale independencia entre los eventos $A_1, \ldots, A_n$ no alcanza con mostrar la independencia de a pares de eventos A_i y A_j.

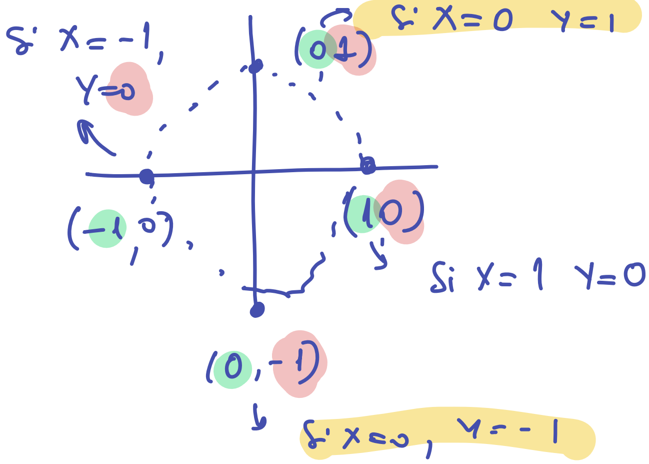
Correlación

$\mathrm{Corr}(X,Y) =$

X

$Y = X^2$





$Z = 0, 1, 2, 3$

con prob $1/4$

$X = \cos\left(\frac{\pi}{2} Z\right)$

$Y = \mathrm{sen}\left(\frac{\pi}{2} Z\right)$

$\mathrm{Corr}(X,Y) = 0$ porque no hay relación lineal entre X e Y pero sí hay una relación no lineal entre X e Y porque $X^2 + Y^2 = 1$.

Otro ejemplo de que $Cov(X, Y) = 0 \not\Rightarrow X$ e Y indep

Sea Z una variable aleatoria discreta que cumple que $P(Z = z) = 0{,}25$ para $z = 0, 1, 2, 3$. Definimos $X = \cos\left(\frac{\pi}{2}Z\right)$ y $Y = \sin\left(\frac{\pi}{2}Z\right)$. Las PMF de X e Y se pueden escribir

$$p_X(x) = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}^T, \quad p_Y(y) = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}^T$$

El producto $p_X(x)p_Y(y)^T$ es

$$p_X(x)p_Y(y)^T = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}^T \cdot \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{bmatrix}$$

Otro ejemplo de que $Cov(X, Y) = 0 \not\Rightarrow X$ e Y indep

$Corr(X, Y) = 0$

Por otro lado, es fácil ver que $E(X) = 0$ y que $E(Y) = 0$.
Como

$$E(XY) = E\left[\cos\left(\frac{\pi}{2}Z\right) \cdot \sin\left(\frac{\pi}{2}Z\right)\right] = E\left[\frac{1}{2}\sin(\pi Z)\right] = 0$$

Entonces $Cov(X, Y) = 0 - 0 \cdot 0 = 0$. Sin embargo, la PMF conjunta de X e Y:

$$p_{X,Y}(x, y) = \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{bmatrix}$$

Entonces, $p_{X,Y}(x, y) \neq p_X(x)p_Y(y)$, a pesar de que $E(XY) = E(X)E(Y)$.