Probabilidad y Estadística

Clase 2

Gonzalo Blanco

2023

# 1. Probabilidad Condicional

Sean A y B dos eventos de un espacio muestral S. Se define la *probabilidad de A condicionado a B* como la probabilidad de que ocurra el evento A **si ocurrió** el evento B. Se lo denota como P(A/B).

Ejemplo: Sea $\varepsilon$ : *tirar un dado y registrar que número sale.*
Con S={1,2,3,4,5,6}, y sean A={2,4,6} y B={1,2,3} dos eventos de interés de S. Calcular P(A/B).

Entonces, debemos determinar la probabilidad de A **si ocurrió B**. Para esto, damos por hecho que al realizar el experimento ocurrió el evento B, es decir salió un 1,2 o 3. Entonces definimos un nuevo espacio muestral S'={1,2,3}, generado por los resultados que hacen que ocurra B, y ahora nos preguntamos, **Si ocurrió B ¿cuál es la probabilidad de que ocurra A?**, o lo que es lo mismo, si ocurrió B ¿ cuál de los posibles resultados que hacen que ocurra B tambíen hacen que ocurra A?. Por definición de probabilidad condicional sabemos que B ocurre por lo tanto los posibles resultados de mi experimento aleatorio son {1,2,3}, y A ocurre si sale un 2,4 o 6. Entonces de los posibles resultados que hacen que ocurra B, solo el 2 hace que tambíen ocurra A, es decir, un solo resultado entre los tres que hacen que ocurra B hace que ocurra A.

$$\Longrightarrow P(A/B) = \frac{1}{3}$$

Si queremos calcular P(A/B) una buena forma de trabajar es definir un nuevo espacio muestral S' cuyos elementos serán aquellos que hagan que ocurra B, es decir serán los elementos de B y a partir de S' determinar la P(A) de forma muy simple, estimando cuantos de los elementos de S' hacen que ocurra A y dividiendo esa cantidad por él numero de elementos de S', como hicimos en el ejemplo anterior.

S'={1,2,3} con #S=3 , la cantidad de elementos de S' que hacen que ocurra A es uno, {2} $\Longrightarrow P(A/B) = 1/3$.

**Recordar que la fórmula $P(A) = \frac{\#A}{\#S}$ es válidad si y solo si el espacio muestral S es equiprobable.**

Ejemplo: $\varepsilon$: extraer sin reemplazo dos pelotitas de una urna con 7 pelotas azules y 4 rojas, donde cada pelotita es distinguible.

Sean los siguientes eventos A= "la primera pelotita es azul" y B="la segunda pelotita es roja".

Determinar P(B/A).

Entonces debemos determinar la probabilidad de B condicionado a A, o lo que es lo mismo la probabilidad de B si ocurre A.

Por definición de probabilidad condicional ocurre A, entonces la primera pelotita que saqué es de color azul. Ahora para que ocurra B la segunda pelotita tiene que ser roja, pero tengamos en cuenta que ocurrió A y que el experimento es sin reemplazo, es decir, saqué la primera pelotita que fue azul y "me la quedé en la mano"(no la volví a poner en la urna), entonces cuando vaya a sacar la 2da pelotita en la urna voy a tener 6 pelotas azules y 4 rojas (ya que tengo una pelotita azul en la mano). Entonces sabiendo ahora que en la urna tengo 6 pelotitas azules y 4 rojas me pregunto ¿Cuál es la probabilidad de sacar una pelotita roja?, o lo que es lo mismo, ¿Cual es la probabilidad de que ocurra B si ocurrió A?. Como estamos en un espacio equiprobable, ya que todas las pelotitas tienen la misma probabilidad de salir entonces tengo 4 pelotitas rojas entre un total de 10 pelotitas $\Longrightarrow P(B/A) = 4/10$.

Veamos ahora la **definición matemática** de probabilidad condicional:

$$\bullet\ P(A/B) = \frac{P(A \cap B)}{P(B)} \quad si\ P(B) \neq 0$$

$$\bullet\ P(B/A) = \frac{P(A \cap B)}{P(A)} \quad si\ P(A) \neq 0$$

**Observación:** Generalmente $P(A/B) \neq P(B/A)$.

En algunos casos esta probabilidad se puede determinar reduciendo el espacio muestral (como en los ejemplos anteriores donde determinamos S') y en otras ocaciones será necesario aplicar la definición anterior.

Ejemplo:

En la provincia de Buenos Aires, el 50 % de la población es mayor de edad, el 30 % hace deporte y el 15 % hace deporte y es mayor de edad. Se elige una persona al azar,

1. Si es mayor de edad, ¿Cuál es la probabilidad de que haga deporte?
2. Si hace deporte, ¿Cuál es la probabilidad de que sea mayor de edad?
3. ¿Cuál es la probabilidad de que sea menor de edad y no haga deporte?

Antes de resolver los puntos 1,2 y 3 identifiquemos la información que tenemos.

El experimento aleatorio es $\varepsilon$: *elegir una persona al azar de la provincia de buenos aires.*, su espacio muestral S viene dado por S=*"cada una de las personas de la provincia de Bs.As"*.

Podemos definir tambíen los siguientes eventos, A=*"es mayor de edad"* $\Longrightarrow P(A) = 0{,}5$
B=*"hace deporte"* $\Longrightarrow P(B) = 0{,}3$
C=*"es mayor de edad y hace deporte"* $\longrightarrow P(C) = 0{,}15$

Ahora sí, resolvamos los incisos 1,2 y 3.

1. Tengo que determinar la probabilidad de que una persona haga deporte si se que es mayor de edad, es decir me pide que calcule la probabilidad de B condicionado a A , P(B/A). Aplico la definición

$P(B/A) = \frac{P(A\cap B)}{P(A)}$ $, con\ P(A) \neq 0$

Notar que $P(A \cap B)$ es la probabilidad de que *una persona sea mayor de edad y haga deporte* es decir es P(C).

$\Longrightarrow P(B/A) = \frac{P(A\cap B)}{P(A)} = \frac{0{,}15}{0{,}5} = 0{,}3$

2. Tengo que determinar la probabilidad de que una persona sea mayor de edad si hace deporte, es decir me pide que calcule la probabilidad de A condicionado a B, P(A/B). Nuevamente aplico la definición

$\Longrightarrow P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{0{,}15}{0{,}3} = 0{,}5$

3. Defino D = *"sea menor de edad y no haga deporte"* a D lo puedo escribir como $D = A^c \cap B^c$ ya que $A^c$= *" sea menor de edad* y $B^c$=*" no haga deporte"*.

Entonces por las leyes de Morgan, vale que,

$P(D) = P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] =$
$= 1 - [0{,}5 + 0{,}3 - 0{,}15] = 0{,}35$

**<u>Axiomas</u>**

Sean A y B dos eventos, vale que:

1. Si P(A)$\geqslant 0$ y $P(A \cap B) \geqslant 0 \quad \Longrightarrow P(B/A) \geqslant 0$
2. $P((A \cup B)/C) = P(A/C) + P(B/C) \quad si\ A \cap B = \phi$, es decir si A y B son mutuamente excluyentes.
3. $P(S/A) = 1$

## 1.1. Teorema de la Multiplicación

Dijimos que $P(A/B) = \frac{P(A\cap B)}{P(B)}$ ,con P(B)$\neq 0$

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

Y lo mismo vale para $P(B/A) = \frac{P(A\cap B)}{P(A)}$ con P(A)$\neq 0$

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

Esto se lo puede generalizar para $n$ eventos y se lo conoce como Teorema de la Multiplicación.

Hagámoslo primero para tres eventos y luego lo generalizamos para $n$. Supongamos que tengo los eventos $A_1, A_2, A_3$ y quiero determinar $P(A_1 \cap A_2 \cap A_3)$,

$P(A_1 \cap A_2 \cap A_3) = P(B \cap A_3)$ , siendo $B = A_1 \cap A_2$

$P(A_1 \cap A_2 \cap A_3) = P(B \cap A_3) = P(B)P(A_3/B)$ , por definición de prob. condicional

$P(A_1 \cap A_2 \cap A_3) = P(B \cap A_3) = P(B)P(A_3/B) = P(A_1 \cap A_2)P(A_3/(A_1 \cap A_2))$

$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2)P(A_3/(A_1 \cap A_2)) = P(A_1)P(A_2/A_1)P(A_3/(A_1 \cap A_2))$, por def. de prob. condicional

$$\Longrightarrow P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2/A_1)P(A_3/(A_1 \cap A_2))$$

Y este resultado se puede generalizar para $n$ eventos, siguiendo el mismo procedimiento,

$$P(A_1 \cap A_2 \cap A_3 \cap ... \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/(A_2 \cap A_1)...P(A_n/(A_1 \cap A_2 \cap A_3 \cap ... \cap A_{n-1}))$$

Ejemplo:

*En una bolsa hay 10 bolitas rojas y 4 azules. Si se sacan 3 bolitas al azar sin reemplazo, ¿Cuál es la probabilidad de sacar 3 bolitas rojas?.*

Defino los eventos $A_i$:*"la i-ésima bolita que saco es roja"*, con i=1, 2, 3.

$\Longrightarrow P(sacar\ 3\ bolitas\ rojas) = P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2/A_1)P(A_3/(A_1 \cap A_2)) = \frac{10}{14}\frac{9}{13}\frac{8}{12}$

Veamos de donde salieron estos números $\frac{10}{14}\frac{9}{13}\frac{8}{12}$

$P(A_1)$ es la probabilidad de que la primera bolita que saco sea roja, como tengo 10 bolitas rojas entre un total de 14, y estamos en un espacio equiprobable, entonces $P(A_1) = \frac{casos\ favorables}{casos\ totales} = \frac{10}{14}$.

$P(A_2/A_1)$ es la probabilidad de que la segunda bolita que saco sea roja **si la primera que saqué también fue roja**. Como el experimento es sin reemplazo, la primera bolita roja no la devolví a la bolsita, entonces cuando vaya a sacar la segunda bolita, en la bolsa habrá 13 pelotitas (en lugar de 14) de las cuales 9 son rojas, entonces como nuevamente estamos en un nuevo espacio muestral equiprobable ya que cualquiera de las 13 bolitas tiene la misma probabilidad de salir, vale que $P(A_2/A_1) = \frac{casos\ favorables}{casos\ totales} = \frac{9}{13}$.

Falta calcular $P(A_3/(A_1 \cap A_2))$, es la probabilidad de que la tercera bolita sea roja **si las 2 anteriores fueron rojas**, vale el mismo razonamiento que arriba, como el experimento es sin reemplazo, ahora en la bolsita tengo 12 pelotitas de las cuales 8 son rojas, como seguimos en un espacio equiprobable entonces $P(A_3/(A_1 \cap A_2)) = \frac{casos\ favorables}{casos\ totales} = \frac{8}{12}$.

## 1.2. Teorema de la Probabilidad Total

Sean $A_1, A_2, A_3, ..., A_n$ eventos de un espacio muestral S, con las siguientes características,

1. $A_1 \cup A_2 \cup A_3 \cup ... \cup A_n = S$
2. $A_i \cap A_j = \phi, \quad \forall\, i \neq j$, es decir que estos eventos $A_i$ son mutuamente excluyentes entre ellos.
3. $P(A_i) > 0, \quad \forall i = 1, 2, 3.., n$

$\Longrightarrow$ decimos que estos eventos $A_1, A_2, A_3, ..., A_n$ son una **partición** de S. Y vale que para cualquier evento $B$ de interés, la probabilidad de $B$ se puede calcular como,

$$P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + P(B/A_3)P(A_3) + ... + P(B/A_n)P(A_n)$$

Demostración:

$B = B \cap S = B \cap (A_1 \cup A_2 \cup A_3 \cup ... \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup ... \cup (B \cap A_n)$, ya que $\cap$ es distriburitva respecto a $\cup$.

Como $A_i \cap A_j = \phi, \ \forall i \neq j \implies (B \cap A_i) \cap (B \cap A_j) = \phi, \ \forall i \neq j$, es decir los eventos $(B \cap A_i)$ $\forall i$ son mutuamente excluyentes entre ellos.

$\implies P(B) = P(B \cap A_1) + P(B \cap A_2) + ... + P(B \cap A_n)$, y si le aplicamos el teorema de la multiplicación o la definiciónd e probabilidad condicional a cada $B \cap A_i$ llego a,

$$P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + ... + P(B/A_n)P(A_n)$$

## 1.3. Teorema de Bayes

Sean $A_1, A_2, ..., A_n$ una partición de un espacio muestral S, y sea B un evento cualquiera de S. Vimos que por el teorema de la probabilidad total, podemos calcular la probabilidad de B como,

$$P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + ... + P(B/A_n)P(A_n)$$

El T. de Bayes nos va a permitir estimar la $P(A_i/B)$, $\forall i$, i=1,2..n.

$P(A_k/B) = \frac{P(A_k \cap B)}{P(B)}$, por definición de prob. condicional

$P(A_k/B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B/A_k)}{P(B)}$, por definiciónd e prob. condicional

$P(A_k/B) = \frac{P(A_k) \cap B}{P(B)} = \frac{P(A_k)P(B/A_k)}{P(B)} = \frac{P(A_k)P(B/A_k)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$, por. T. de la probabilidad total

$$\implies P(A_k/B) = \frac{P(A_k)P(B/A_k)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$$

Ejemplo:

*Supongamos que queremos estudiar la probabilidad de estar infectados por una enfermedad H en la Argentina. Sean los eventos,*

$B_1$= tener H

$B_2$= no tener H

*Para esto se testiaron a un grupo de control que puede considerarse representativo de todo el país y se obtuvo que el 0,1 % de la población está enfermo, es decir*

$P(B_1) = 0{,}001$, *esto se traduce como la probabilidad de que una persona esté infectada con H es del 0,1 %.*

$P(B_2) = 0{,}999$, *esto se traduce como la probabilidad de que una persona no esté infectada con H es del 99,9 %.*

*Pero, los desarolladores de los test informaron que si la persona está enferma hay un 98 % de probabilidad de que el test de positivo, y si la persona está sana hay un 3 % de probabilidad de que el test de un falso positivo.*

*Si elegimos una persona al azar y el test le da positivo, ¿ Cuál es la probabilidad de que esté enfermo?*

Primero, como siempre, organicemos la información que tenemos. Además de los eventos $B_1$ y $B_2$ , podemos definir los eventos $\oplus$ = *el test da positivo* y el evento $\ominus$ = *el test da negativo*. Por la información que brindaron los desarrolladores del test sabemos que,

$P(\oplus/B_1) = 0,98$, la probabilidad de que el test de positivo si la persona está enferma con H.

$P(\oplus/B_2) = 0,03$, la probabilidad de que el test de positivo si la persona está sana.

A partir de estos dos datos podemos deducir que,

$P(\ominus/B_1) = 1 - P(\oplus/B_1) = 0,02$, la probabilidad de que el test de negativo si la persona está enferma con H.

$P(\ominus/B_2) = 1 - P(\oplus/B_2) = 0,97$, la probabilidad de que el test de negativo si la persona está sana.

Y nos interesa determinar la probabilidad de que una persona, a la cual el test le da positivo, esté enferma. Es decir, necesitamos determinar $P(B_1/\oplus)$. Y esto lo podemos resolver aplicando el T. de Bayes ya que $B_1, B_2$ son una partición de S,

$$\Longrightarrow P(B_1/\oplus) = \frac{P(\oplus/B_1)P(B_1)}{P(\oplus/B_1)P(B_1) + P(\oplus/B_2)P(B_2)} = \frac{0{,}98x0{,}001}{0{,}98x0{,}001 + 0{,}03x0{,}999} = 0{,}032 = 3,2\,\%$$

La probabilidad de estar enfermo, si el test me da positivo, es de un 3,2 %.

Ejemplo:

*Las máquinas A,B y C producen el 70 %, 20 % y 10 %, respectivamente, de los productos de una fábrica. Además, se sabe que estas máquinas no son perfectas y del total de productos producido por cada máquina el 2 %, 4 % y 6 %, respectivamente, son defectuosos.*

*Si agarro un producto al azar de la fábrica,*

1. *¿Cuál es la probabilidad de que el producto sea defectuoso?*
2. *Si el producto seleccionado es defectuoso, ¿Cuál es la probabilidad de que haya sido producido por la máquina B?*

Primero organicemos la información.

El experimento aleatorio sería $\varepsilon$: *tomar un producto al azar de la fábrica.*

Podemos definir los siguientes eventos,

*A=producido en la máquina A*, con P(A)=0.7.

*B=producido en la máquina B*, con P(B)=0.2.

*C=producido en la máquina C*, con P(C)=0.1.

*D=producto defectuoso*, desconocemos P(D).

*P(D/A)=0.02*, probabilidad de que el producto sea defectuoso si fue producido en la máquina A.

*P(D/B)=0.04*, probabilidad de que el producto sea defectuoso si fue producido en la máquina B.

*P(D/C)=0.06*, probabilidad de que el producto sea defectuoso si fue producido en la máquina C.

1. Queremos determinar P(D)

Como A,B,C conforman una partición de S, entonces podemos aplicar el T. de la probabilidad total para calcular P(D),

$P(D) = P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) = 0{,}02*0{,}7 + 0{,}04*0{,}2 + 0{,}06*0{,}1 = 0{,}028 = 2{,}8\,\%$.

La probabilidad de que un producto sea defectuoso es del 2.8 %.

2. Si el producto seleccionado es defectuoso, ¿Cuál es la probabilidad de que haya sido producido por la máquina B?. Nos está pidiendo que calculemos P(B/D).

Nuevamente como A,B,C son una partición de S, podemos aplicar el T. de Bayes para determinar P(B/D),

$P(B/D) = \frac{P(B)P(D/B)}{P(A)P(D/A)+P(B)P(D/B)+P(C)P(D/C)} = 0{,}28 = 28\,\%$

La probabilidad de que un producto que es defectuoso provenga de la máquina B es del 28 %.

## 1.4. Independencia estadística

Sean A y B dos eventos. Se dice que son independientes si,

$$P(A/B) = P(A) \quad o \quad P(B/A) = P(B)$$

Esto significa que la ocurrencia de un evento no condiciona o modifica la probabilidad de ocurrencia del otro.

Ejemplo:

Tengo un mazo de 40 cartas españolas y saco una carta al azar, sean los eventos

*A= 1 de cualquier palo*

*B= cualquiera de espadas*

¿Son A y B eventos independientes?

Para que lo sean se debe cumplir que P(A/B)=P(A), pero antes de calcularlo organicemos la información

$\varepsilon$ :*sacar una carta al azar de un mazo de 40 cartas españolas.*

*S*={*cada una de las cartas del mazo*}, con #S=40.

También sabemos que,

#A=4

#B=10

Como todas las cartas tienen la misma probabildid de salir, entonces estamos en un espacio muestral equiprobable

$\Longrightarrow P(A) = \frac{\#A}{\#S} = \frac{casos\ favorables}{casos\ totales} = \frac{4}{40} = \frac{1}{10}$

Veamos si P(A/B)=P(A),

$P(A/B) = \frac{P(A\cap B)}{P(B)}$ Calculemos $P(A \cap B)$ y $P(B)$,

$A \cap B$ ={*sacar un uno y que sea de espadas*}={*sacar el 1 de espadas*}, con $\#(A \cap B) = 1$

$\Longrightarrow P(A \cap B) = \frac{\#(A\cap B)}{\#S} = \frac{1}{40}$

$P(B) = \frac{\#B}{\#S} = \frac{10}{40}$

$\Longrightarrow P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{1/40}{10/40} = \frac{1}{10} = P(A)$

Entonces A y B son eventos independientes.