# Probabilidad y Estadística Fundamental

## Probabilidad condicional. Independencia y regla de Bayes.

Profesor: Nicolás López

Universidad Nacional de Colombia

# Contenido

# Contenido

# Introducción

Si desea calcular la probabilidad de ocurrencia de:

1. Daltonismo condicionando por el sexo del paciente.
2. Muerte por cáncer pulmonar para un fumador.
3. Preferencia de un juguete dependiendo de su color.
4. ...

De cualquier forma, esta información parcial conocida acerca del experimento aleatorio antes de conocer su resultado final *podría* cambiar la estructura probabilística de los posibles resultados.

# Probabilidad condicional

## Apuesta al lanzar dos dados

- Suponga que al lanzar dos dados corrientes (uno rojo, otro verde) de 6 caras le proponen apostar que al menos uno de los resultados sea 6.
- Como decide declinar la propuesta por la baja probabilidad de ocurrencia del evento, al arrojar los dados, le dicen que los resultados obtenidos fueron diferentes. ¿Decidiría volver a la apuesta?.
- Tenga en cuenta. El conjunto de posibles resultados de un experimento aleatorio puede verse restringido a un subconjunto $B$ por el conocimiento de información previa respecto al mismo.

# Probabilidad condicional

## Apuesta al lanzar dos dados

### Ejemplo. Lanzamiento de dos dados corrientes (uno rojo, otro verde) de 6 caras

Experimento aleatorio Lanzamiento de dos dados corrientes (uno rojo, otro verde) de 6 caras.

Espacio muestral El espacio muestral $\Omega$ está conformado por las duplas:

$$\begin{array}{cccccc}
(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\
(2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\
(3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\
(4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\
(5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\
(6,1), & (5,6), & (6,3), & (6,4), & (6,5), & (6,6)
\end{array}$$

¿A qué es igual la probabilidad que al menos uno de los resultados sea 6?

## Ejemplo. Lanzamiento de dos dados **corrientes** (uno rojo, otro verde) de 6 caras

Experimento aleatorio Lanzamiento de dos dados corrientes (uno rojo, otro verde) de 6 caras.

Espacio muestral El espacio muestral $\Omega$ está conformado por las duplas:

$$\begin{array}{cccccc}
(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\
(2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\
(3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\
(4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\
(5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\
(6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6)
\end{array}$$

El exp. es laplaciano, entonces, si $A$ es el evento de interés, $P(A) = 11/36$.

Suponga ahora que conoce que los resultados fueron diferentes, es decir, los resultados $(1,1)$, ..., $(6,6)$ no fueron observados, ¿Cambia la estructura probabilística del experimento? ¿la probabilidad que $A$ (obtener al menos un numero seis) sigue siendo igual con esta información adicional?

# Probabilidad condicional

## Apuesta al lanzar dos dados

### Ejemplo. Lanzamiento de dos dados **corrientes** (uno rojo, otro verde) de 6 caras

El "nuevo" espacio muestral (llamémoslo $B$) está conformado por las duplas:

| | | | | | |
|---|---|---|---|---|---|
| | (1,2), | (1,3), | (1,4), | (1,5), | (1,6), |
| (2,1), | | (2,3), | (2,4), | (2,5), | (2,6), |
| (3,1), | (3,2), | | (3,4), | (3,5), | (3,6), |
| (4,1), | (4,2), | (4,3), | | (4,5), | (4,6), |
| (5,1), | (5,2), | (5,3), | (5,4), | | (5,6), |
| (6,1), | (5,6), | (6,3), | (6,4), | (6,5), | |

Ahora, ¿a qué es igual la probabilidad que al menos uno de los resultados sea 6?

# Probabilidad condicional

Apuesta al lanzar dos dados

## Ejemplo. Lanzamiento de dos dados **corrientes** (uno rojo, otro verde) de 6 caras

El "nuevo" espacio muestral $B$ está conformado por las duplas:

| | | | | | |
|---|---|---|---|---|---|
| | (1,2), | (1,3), | (1,4), | (1,5), | (1,6), |
| (2,1), | | (2,3), | (2,4), | (2,5), | (2,6), |
| (3,1), | (3,2), | | (3,4), | (3,5), | (3,6), |
| (4,1), | (4,2), | (4,3), | | (4,5), | (4,6), |
| (5,1), | (5,2), | (5,3), | (5,4), | | (5,6), |
| (6,1), | (6,2), | (6,3), | (6,4), | (6,5), | |

El exp. es laplaciano, entonces, si $A$ es el evento de interés, ¿$P(A) = 10/30 = 0{,}333$? ó ¿$P(A) = 11/36 = 0,305$?.

# Probabilidad condicional

## Apuesta al lanzar dos dados

Note que $11/36$ no utiliza el conocimiento previo acerca del evento *Observar resultados diferentes*, mientras que $10/30$ si. Sea $B$ este evento, entonces $10/30$ es la probabilidad de $A$ **suponiendo que el evento B sucedió** (e.d. el "nuevo" espacio es $B$), así:

$$P(A) = 11/36$$

Es diferente a

$$\mathbf{P(A|B) = 10/30}$$

# Probabilidad condicional

## Definición

La frecuencia relativa condicional está dada por

$$f_{A|B} = \frac{n_{A \cap B}}{n_B}$$

Al repetir el experimento aleatorio $N$ veces, $f_{A|B}$ corresponde a la proporción del total de eventos en $B$ que además resultaron en $A$. Para el ejemplo anterior:

1. $A =$ Obtener al menos un numero seis al lanzar los dos dados.
2. $B =$ Obtener resultados diferentes al lanzar los dos dados.
3. $A|B =$ Obtener al menos un numero seis **dado que** los resultados son diferentes al lanzar los dos dados.

# Probabilidad condicional

## Definición

Si lanzamos el par de dados $N = 36$ veces se tiene que:

- Como $P(B) = 30/36$, aproximadamente en $n_B = 30$ de estas repeticiones se obtendrán resultados diferentes para cada dado.
- Como $P(A \cap B) = 10/36$, aproximadamente $n_{A \cap B} = 10$ de estas repeticiones tendrán al menos un seis y los resultados serán diferentes entre si.

Se obtiene entonces la frecuencia relativa condicional como

$$f_{A|B} = \frac{n_{A \cap B}}{n_B} \approx \frac{10}{30} = P(A|B)$$

# Probabilidad condicional

## Definición

Desde la noción de regularidad estadística, obsérvese que

$$f_{A|B} = \frac{n_{A\cap B}}{n_B} = \frac{\frac{n_{A\cap B}}{N}}{\frac{n_B}{N}} = \frac{f_{A\cap B}}{f_B}$$

Dónde $A$ y $B$ son eventos, por lo cual:

$$f_A = \frac{n_A}{N} \xrightarrow[N\to\infty]{} P(A)$$

Y además

$$f_{A\cap B} = \frac{n_{A\cap B}}{N} \xrightarrow[N\to\infty]{} P(A\cap B)$$

Lo cual motiva la siguiente definición.

# Probabilidad condicional



## Probabilidad condicional

Sean $A$ y $B$ dos eventos. Se define la probabilidad del evento $A$ bajo la condición $B$ como

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

Esta última igualdad al multiplicar por $P(B)$ en ambos lados puede verse como

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

# Probabilidad condicional

## Definición

Como comentarios

- Note que $P(B)$ debe ser mayor de 0 para la división numérica en la definición (y si fuera 0, desde el punto de vista de la frecuencia relativa, sería un evento que nunca sucedería).
- La probabilidad condicional cumple
  1. $P(A|B) \geq 0$ para cualquier evento $A$.
  2. Si $A = E_1 \cup E_2 \cup ... \cup E_m$ con $E_1$, $E_2$,..., $E_m$ disyuntos entre si, entonces

$$P(A|B) = P(E_1|B) + P(E_2|B) + ... + P(E_m|B)$$

  3. Condicionado a $B$, la probabilidad de $B$ es 1, es decir, $P(B|B) = 1$.

Así, $P(A|B)$ es una medida de probabilidad.

## Ejemplo. Grupo sanguíneo

Se tienen 1000 personas en una población. 600 hombres (de los cuales 200 tienen el tipo de sangre A) y 400 mujeres (de las cuales 180 tienen el tipo de sangre A)

- Defina los eventos mencionados en el enunciado.
- ¿Cuál es la probabilidad de que una persona elegida al azar tenga el tipo de sangre A?
- ¿Cuál es la probabilidad de que un hombre elegida al azar tenga el tipo de sangre A?
- ¿Cuál es la probabilidad de que una mujer elegida al azar tenga el tipo de sangre A?

## Ejemplo. Grupo sanguíneo

- Defina los eventos mencionados en el enunciado. $A =$ Persona seleccionada de tipo de sangre A. $H =$ Persona seleccionada es hombre. $M =$ Persona seleccionada es mujer.
- ¿Cuál es la probabilidad de que una persona elegida al azar tenga el tipo de sangre A?: $P(A) = (200 + 180)/1000$
- ¿Cuál es la probabilidad de que un hombre elegido al azar tenga el tipo de sangre A?: $P(A|H) = 200/600$, igual a:

$$P(A|H) = \frac{P(A \cap H)}{P(H)} = \frac{200/1000}{600/1000} = \frac{200}{600}$$

- ¿Cuál es la probabilidad de que una mujer elegida al azar tenga el tipo de sangre A? $P(A|M) = 180/400$

$$P(A|M) = \frac{P(A \cap M)}{P(M)} = \frac{180/1000}{400/1000} = \frac{180}{400}$$

# Eventos independientes

## Definición

Algunas veces, el conocimiento de $B$ no afecta la probabilidad de ocurrencia de un evento $A$, es decir, $P(A|B) = P(A)$, con lo cual

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

### Eventos independientes

Sean $A$ y $B$ dos eventos. Se dice que los eventos son independientes si y solo si

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

Si la igualdad no se cumple, los eventos son **dependientes**.

## Ejemplo. Hipertensos y fumadores

Se sabe que en una población el chance de ser hipertenso es 0.2, mientras que el de ser fumador es 0.45. Además, la mitad de los individuos es hipertenso o fumador. ¿Ser fumador es independiente de ser hipertenso en esta población?

# Eventos independientes

## Ejemplo

### Ejemplo. Hipertensos y fumadores

Sea

$$H = \text{Individuo hipertenso de la pob.}$$

$$F = \text{Individuo fumador de la pob.}$$

Si hubiera independencia, se tendría que $P(H \cap F) = P(H)P(F)$, pero como

$$P(H)P(F) = 0{,}2 \times 0{,}4 = 0{,}09$$

$$P(H \cap F) = P(H) + P(F) - P(H \cup F) = 0{,}2 + 0{,}45 - 0{,}5 = 0{,}15$$

los eventos no son independientes. En efecto, se nota que:

$$P(H|F) = \frac{0{,}15}{0{,}45} = 0{,}33$$

Mientras que $P(H) = 0,2$. Así, ser fumador aumenta la probabilidad de ser hipertenso.

# Probabilidad total

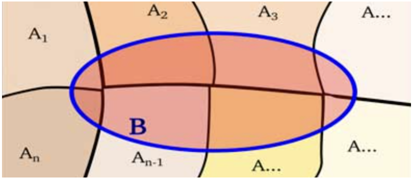
## Definición

Al contar con una **partición** $A_1, ..., A_n$ sobre $\Omega$:

1. Exhaustiva: $\cup A_i = \Omega$.
2. Mutuamente excluyente: $A_i \cap A_j = \emptyset$, $i \neq j$.

¿A qué es igual la probabilidad de cualquier evento $B$, inducida por la partición?

Figura 1: Evento $B$ inducido por una partición

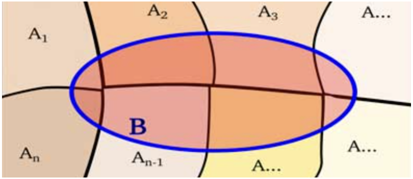


# Probabilidad total

## Definición

Suponga que $A_1, ..., A_n$ son las $n$ causas posibles de que el evento $B$ suceda (o las $n$ ubicaciones posibles en las que el evento $B$ puede ocurrir). El **teorema de probabilidad total** muestra la probabilidad de $B$ como la suma de las probabilidades del evento $B$ bajo todas las causas (o lugares) posibles $A_1, ..., A_n$.

Figura 2: Evento $B$ inducido por una partición



## Probabilidad total

Sea $B$ un evento cualquiera y $A_1, ..., A_n$ una partición sobre $\Omega$:

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

Donde $P(A_i)$ se conocen como las probabilidades *apriori* (antes de que suceda $B$) de la partición. Se tiene además que $P(A_j|B)$ está dado por:

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

Son las probabilidades *aposteriori* (después de que suceda B) de la partición.

## Ejemplo. Virus en dos zonas de la ciudad

Una ciudad está dividida en 2 zonas de salud: norte y sur, que atienden al 40 % y el 60 % de la población, respectivamente. Un estudio llevado a cabo en ambas zonas de salud revela que un 30 % de los individuos de la zona norte y un 40 % de los de la zona sur fueron infectados por covid. ¿Cuál es la probabilidad de que un individuo elegido al azar en esa ciudad hubiera tenido dicho virus?

## Ejemplo. Virus en dos zonas de la ciudad

En este caso, el evento $B$ representa que el individuo tuvo el virus, mientras las partición dónde el evento puede suceder es: $A_1 =$ Individuo originario de la zona norte, y $A_2 =$ Individuo originario de la zona sur. Se pregunta $P(B)$:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) = 0{,}3 \times 0{,}4 + 0{,}4 \times 0{,}6 = 0{,}36$$

Con lo cual, hay un chance de 0.36 de que un individuo seleccionado al azar tenga el virus, ó, de manera equivalente, un 36 % de la población fue infectada por el virus.

# Teorema de Bayes

## Definición

En las probabilidades *aposteriori* presentadas, se sabe que

$$P(A_j \cap B) = P(B|A_j)P(A_j)$$

Y además

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

Con lo cual se tiene el teorema de Bayes como sigue:

### Teorema de Bayes

Sea $B$ un evento cualquiera y $A_1, ..., A_n$ una partición sobre $\Omega$. El teorema de Bayes indica que

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

Para $j = 1, ..., n$.

## Ejemplo. Virus en dos zonas de la ciudad (cont.)

Si un habitante de la ciudad fue infectado por el coronavirus de estudio, ¿cuál es la probabilidad que provenga de la zona norte? ¿cuál es la probabilidad que provenga de la zona sur?

## Ejemplo. Virus en dos zonas de la ciudad (cont.)

Se preguntan las probabilidades *aposteriori* de $P(A_1)$, con $P(A_1) = 0,40$, y de $P(A_2)$, con $P(A_2) = 0,60$. Es decir:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{2} P(B|A_i)P(A_i)} = \frac{0,3 \times 0,4}{0,36} = 0{,}33$$

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{\sum_{i=1}^{2} P(B|A_i)P(A_i)} = \frac{0,4 \times 0,6}{0,36} = 0{,}66$$

Con lo cual, si un individuo tuvo el virus, es dos veces más probable que provenga de la zona sur que de la zona norte. Note además que la probabilidad de venir de la zona norte disminuye de un $40\,\% = P(A_1)$ a un $33\,\% = P(A_1|B)$, con el conocimiento de la pre existencia del virus.

# Conclusiones

- La probabilidad de un evento puede cambiar al ser condicionada, múltiples ejemplos son observados en la vida diaria: chance diferencial de ocurrencia de un evento determinado entre razas, nivel educativo, hábitos (fumador, deportista, sedentario), entre otros.
- La probabilidad de un evento determinado se puede ver como la suma de probabilidades del evento condicionada bajo una partición determinada.