Tema 4
Probabilidad condicionada: teoremas básicos. Independencia de sucesos

1. Probabilidad condicionada. Espacio de probabilidad condicionado

La probabilidad condicionada es uno de los conceptos clave en Teoría de la Probabilidad. En el tema anterior se ha introducido el concepto de probabilidad considerando que la única información sobre el experimento era el espacio muestral. Sin embargo, hay situaciones en las que se incorpora información suplementaria como puede ser que ha ocurrido otro suceso, con lo que puede variar el espacio de resultados posibles y consecuentemente, sus probabilidades. En este contexto aparece el concepto de probabilidad condicionada.

El objetivo es analizar cómo afecta el conocimiento de la realización de un determinado suceso a la probabilidad de que ocurra cualquier otro.

La probabilidad condicionada tiene una clara interpretación en espacios muestrales finitos en los que puede aplicarse la regla de Laplace.

Definición.- Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico arbitrario y A un suceso ($A \in \mathcal{A}$) tal que $P(A) > 0$. Para cualquier otro suceso $B \in \mathcal{A}$, se define la **probabilidad condicionada de B dado A** o **probabilidad de B condicionada a A** como

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

Observemos que la condición $P(A) > 0$ es necesaria para que la definición tenga sentido. Por otra parte, la idea intuitiva de probabilidad condicionada hace lógica esta restricción ya que si $P(A) = 0$, A es un suceso imposible y no tiene sentido condicionar a él.

Notemos que, sabiendo que $A \in \mathcal{A}$ ha ocurrido, tenemos una nueva evaluación de la probabilidad de cada suceso ($P(B) \longrightarrow P(B/A)$), o sea, tenemos una nueva función de conjunto sobre $(\Omega, \mathcal{A})$. Probamos a continuación que, efectivamente, esta función es una medida de probabilidad sobre $(\Omega, \mathcal{A})$

Teorema 1

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico y sea un suceso $A \in \mathcal{A}$, tal que $P(A) > 0$. Entonces $(\Omega, \mathcal{A}, P(\cdot/A))$, en donde $P(\cdot/A)$ es la definida anteriormente, es un espacio probabilístico.

Demostración.- Basta probar que $P(\cdot/A)$ es una probabilidad. Evidentemente,

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \geq 0 \qquad \forall B \in \mathcal{A}.$$

También,

$$P(\Omega/A) = \frac{P(\Omega \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

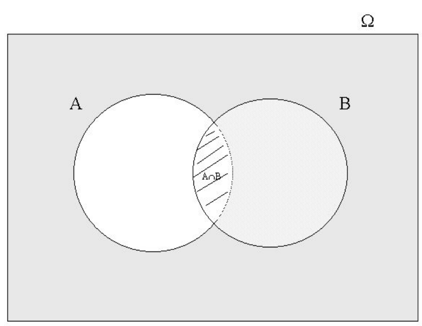
Por último, si $\{A_n\}_{n\in\mathbb{N}}$ es una colección de conjuntos disjuntos de $\mathcal{A}$, entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n/A\right) = \frac{P\left[\left(\bigcup_{n=1}^{\infty} A_n\right) \cap A\right]}{P(A)} = \frac{P\left[\bigcup_{n=1}^{\infty} (A_n \cap A)\right]}{P(A)} =$$

$$\frac{\sum_{n=1}^{\infty} P(A_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} P(A_n/A)$$

Nota: Al condicionar a un suceso $A \in \mathcal{A}$, con $P(A) > 0$, los sucesos de interés en el experimento son sólo aquellos que tienen intersección no vacía con A, ya que si B es tal que $B \cap A = \emptyset$, entonces $P(B/A) = 0$. Además por la propia definición

$$\forall B \in \mathcal{A}, \quad P(B/A) = P(B \cap A/A)$$



O sea, en realidad, estamos haciendo una transformación del espacio muestral, pasando de Ω a A, ya que si A ha ocurrido, no puede haber ocurrido ningún resultado elemental de Ω que no esté en A.
Esto nos lleva a definir un nuevo espacio probabilístico con espacio muestral A, como probamos a continuación, que se denomina **espacio de probabilidad condicionado**

Teorema 2

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico y $A \in \mathcal{A}$ tal que $P(A) > 0$. Consideramos la clase de conjuntos

$$\mathcal{A}_A = \mathcal{A} \cap A = \{B \cap A \ / \ B \in \mathcal{A}\} \quad (\subset \mathcal{P}(A))$$

y la función

$$P_A : \mathcal{A}_A \longrightarrow \mathbb{R}$$

dada por $P_A(C) = \dfrac{P(C)}{P(A)}$ que está bien definida ya que $C \in \mathcal{A}$.
Entonces

1. $\mathcal{A}_A$ es una σ-álgebra contenida en $\mathcal{A}$ (con espacio total A).

2. P_A es una medida de probabilidad sobre $\mathcal{A}_A$

En definitiva, $(A, \mathcal{A}_A, P_A)$ es un espacio probabilístico.

Demostración
1) Evidentemente $\mathcal{A}_A \subset \mathcal{A}$. Veamos que es una σ-álgebra:

1) Si $C \in \mathcal{A}_A$ y si representamos por C^* el complementario de C en A y por $\overline{C}$ el complementario de C en Ω, $C^* = \overline{C} \cap A$ y por ser $\mathcal{A}$ una σ-álgebra, $C^* \in \mathcal{A}_A$.

2) Sea $\{C_n\}_{n \in \mathbb{N}} \in \mathcal{A}_A$, dado que $C_n = B_n \cap A$ con $B_n \in \mathcal{A}$

$$\bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} [B_n \cap A] = \left[\bigcup_{n=1}^{\infty} B_n\right] \cap A \in \mathcal{A}_A$$

por ser $\mathcal{A}$ una σ-álgebra ($\bigcup_{n=1}^{\infty} B_N \in \mathcal{A}$).

2) Veamos que P_A es una medida de probabilidad

- $P_A(C) \geq 0 \quad \forall C \in \mathcal{A}_A$

- $P_A(A) = \dfrac{P(A)}{P(A)} = 1$

- Dados $\{C_n\}_{n \in \mathbb{N}}$ incompatibles o disjuntos dos

$$P_A\left(\bigcup_n C_n\right) = \frac{P\left(\bigcup_n C_n\right)}{P(A)} = \frac{\sum_n P(C_n)}{P(A)} = \sum_n P_A(C_n)$$

Al espacio $(A, \mathcal{A}_A, P_A)$ se le denomina **espacio de probabilidad condicionado**.

Notemos que los espacios de probabilidad $(\Omega, \mathcal{A}, P(\cdot/A))$y $(A, \mathcal{A}_A, P_A)$ son equivalentes en el sentido de que las medidas de probabilidad están determinadas una por otra a través de las relaciones

$$\forall B \in \mathcal{A} \quad P(B/A) = P_A(A \cap B)$$

$$\forall C \in \mathcal{A}_A \quad P_A(C) = \frac{P(C)}{P(A)} = \frac{P(C \cap A)}{P(A)} = P(C/A)$$

1. Teoremas básicos de probabilidad condicionada

La probabilidad de la intersección de dos sucesos se puede deducir directamente de la definición de probabilidad condicionada y se obtiene como

$$P(A \cap B) = P(A)P(B/A) \quad \text{si } P(A) > 0$$

o bien

$$P(A \cap B) = P(B)P(A/B) \quad \text{si } P(B) > 0$$

Si uno de los dos tiene probabilidad nula, la probabilidad condicionada a él no tiene sentido. Si los dos tienen probabilidad nula, entonces la probabilidad de la intersección es evidentemente cero, pero no puede expresarse en función de las probabilidades condicionadas puesto que éstas no existen.

Estas expresiones se generalizan, mediante el teorema de la probabilidad compuesta o regla de la multiplicación, al cálculo de la probabilidad de la intersección de más de dos sucesos que se producen concatenadamente.

Teorema de la probabilidad compuesta o Regla de multiplicación

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $A_1, A_2, \ldots, A_n \in \mathcal{A}$ con $P\left[\bigcap_{i=1}^{n-1} A_i\right] > 0$, entonces

$$P\left[A_1 \cap A_2 \cap \cdots \cap A_n\right] = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdot \ldots \cdot P\left[A_n/A_1 \cap A_2 \cap \cdots \cap A_{n-1}\right]$$

Demostración.- Es claro que $A_1 \cap \cdots \cap A_{n-1} \subseteq A_1 \cap \cdots \cap A_{n-2} \subseteq A_1 \cap A_2 \cap A_3 \subseteq A_1 \cap A_2 \subseteq A_1$. Por tanto, si el primero tiene probabilidad positiva, las restantes también, y todas las probabilidades condicionadas tienen sentido.

La demostración se hace por inducción. Para $n = 2$ es la regla de la multiplicación dada por la definición de probabilidad condicionada. Suponemos que la expresión es cierta para la intersección de $n - 1$ sucesos. Entonces

$$P(A_1 \cap \cdots \cap A_n) = P\left((A_1 \cap \cdots \cap A_{n-1}) \cap A_n\right) = P(A_1 \cap \cdots \cap A_{n-1})P(A_n/A_1 \cap \cdots \cap A_{n-1})$$

y ahora se aplica la hipótesis de inducción a la primera probabilidad obteniendose el resultado deseado.

Nota: Este resultado es especialmente útil en experimentos compuestos de varias etapas en los que las probabilidades de los sucesos en cada etapa dependen de los resultados obtenidos en las anteriores.

Ejemplo.- Se extraen sucesivamente, y sin reemplazamiento, tres bolas de una urna que contiene 7 bolas blancas y tres negras. Calcular la probabilidad de que las dos primeras bolas extradas sean blancas y la tercera negra.

El experimento consta de tres etapas y, al no devolverse la bola extraída de la urna en cada etapa, la probabilidad de los resultados que pueden darse en las extracciones sucesivas depende del resultado en la anterior.

Si consideramos los sucesos

- B_1: Salir bola blanca en la primera extracción,
- B_2: Salir bola blanca en la segunda extracción,
- N_3: Salir bola negra en la tercera extracción,

la probabilidad que nos piden es $P(B_1 \cap B_2 \cap N_3)$ que, aplicando la regla de multiplicación, se calcula de la siguinete forma:

$$P(B_1 \cap B_2 \cap N_3) = P(B_1)\, P(B_2/B_1)\, P(N_3/B_1 \cap B_2) = \frac{7}{10}\,\frac{6}{9}\,\frac{3}{8}$$

Teorema de la probabilidad total

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ un sistema completo de sucesos o partición de Ω con $P(A_n) > 0,\ \forall n \in \mathbb{N}$. Sea B un suceso cualquiera de $\mathcal{A}$, entonces

$$P(B) = \sum_{n=1}^{\infty} P(B/A_n)P(A_n).$$

Demostración.- En efecto, B se puede escribir como una unión disjunta de la forma

$$B = B \cap \Omega = B \cap \left(\bigcup_{n=1}^{\infty} A_n \right) = \bigcup_{n=1}^{\infty} (B \cap A_n)$$

y por la propiedad de aditividad numerable de la probabilidad

$$P(B) = \sum_{n=1}^{\infty} P(B \cap A_n).$$

Ahora, aplicando el Teorema de la probabilidad compuesta se obtiene el resultado deseado

$$P(B) = \sum_{n=1}^{\infty} P(B/A_n)P(A_n).$$

Interpretación.- Los sucesos A_n pueden interpretarse como las distintas causas (o circunstancias) por las que puede ocurrir el suceso B. Entonces el teorema de la probabilidad total viene a decir que si el suceso B puede ocurrir por alguna de las causas A_n, la probabilidad de que ocurra es la suma de las probabilidades de las causas ($P(A_n)$) por la probabilidad del suceso B condicionado a la causa ($P(B/A_n)$).

Ejemplo 1.- Se tienen dos urnas: la urna 1 contiene 2 bolas blancas y 2 negras. La urna 2 tiene dos bolas blancas y 3 negras (todas distinguibles). Se elige una urna al azar y se extrae una bola. ¿Cuál es la probabilidad de que sea blanca?

Consideramos los sucesos

$$\left.\begin{array}{ll} A_1: \text{ elegir la urna 1} & P(A_1) = 1/2 \\ \\ A_2: \text{ elegir la urna 2} & P(A_2) = 1/2 \end{array}\right\} A_1 \cap A_2 = \emptyset \text{ y exhaustivos}$$

y sea $B:$ extraer una bola blanca

$$P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) = \frac{2}{4}\frac{1}{2} + \frac{2}{5}\frac{1}{2} = \frac{36}{80} = \frac{9}{20}$$

Ejemplo 2.- Un determinado tipo de batería se produce en tres plantas diferentes con distintos equipos y operaciones. Supóngase que los promedios semanales del número de baterías producidas en cada una de estas tres plantas son 500, 2000 y 1500, respectivamente. Además, supongamos que las probabilidades de producir una batería defectuosa en cada una de las tres plantas son respectivamente 0.020, 0.015 y 0.030.
Si se supone que las baterías producidas por las tres plantas se destinan a un fabricante de automóviles que recibe semanalmente 4000 baterías. ¿Cuál es la probabilidad de que la batería instalada por el fabricante en un automóvil sea defectuosa?

Definimos los sucesos

$$\begin{array}{l} D: \text{ la batería defectuosa} \\ E_i: \text{ la batería se ha fabricado en la planta } E_i, \quad i = 1, 2, 3 \end{array}$$

La probabilidad de que una batería seleccionada al azar haya sido fabricada en cada una de las fábricas es

$$P(E_1) = \frac{500}{4000}, \qquad P(E_2) = \frac{2000}{4000}, \qquad P(E_3) = \frac{1500}{4000}$$

Además

$$P(D/E_1) = 0.020, \qquad P(D/E_2) = 0.015, \qquad P(D/E_3) = 0.030$$

Si aplicamos el teorema anterior

$$P(D) = \sum_{i=1}^{3} P(D/E_i)P(E_i) = \frac{500}{4000}0.020 + \frac{2000}{4000}0.015 + \frac{1500}{4000}0.030 = 0.02125$$

Teorema de Bayes o de la probabilidad inversa

En las mismas condiciones del Teorema de la probabilidad total

$$P(A_n/B) = \frac{P(B/A_n)P(A_n)}{\sum_{n\in\mathbb{N}} P(B/A_n)P(A_n)}$$

Demostración.- Por la definición de probabilidad condicionada y aplicando el Teorema de la probabilidad compuesta

$$P(A_n/B) = \frac{P(B/A_n)P(A_n)}{P(B)}$$

y aplicando el Teorema de la probabilidad total en el denominador se obtiene el resultado deseado.

El razonamiento lógico que subyace en el cálculo de estas probabilidades es el siguiente: Interpretar, de nuevo, el suceso B como el resultado obtenido al realizar un experimento y los sucesos A_n como el conjunto de todas las "causas" que pueden producir la aparición del suceso B; entonces, si para cada "causa" conocemos su *probabilidad a priori* $P(A_n)$ y la *verosimilitud* $P(B/A_n)$ de que el suceso B haya sido causado por A_n, la ocurrencia de B, nos permite asignar, mediante la aplicación del Teorema de Bayes, una "probabilidad a posteriori" $P(A_n/B)$ al suceso de que la verdadera causa haya sido A_n.

Ejemplo 1.- Se tienen dos urnas: la urna 1 contiene 3 bolas blancas y 2 negras. La urna 2 tiene dos bolas blancas y 3 negras (todas distinguibles). Se elige una urna al azar y se extrae una bola. Si la bola resulta ser blanca. ¿Cuál es la probabilidad de que proceda de la urna 1? ¿y de la 2?

$$P(A_1/B) = \frac{P(B/A_1)P(A_1)}{P(B)} = \frac{\frac{3}{5}\frac{1}{2}}{\frac{1}{2}} = 3/5$$

$$P(A_2/B) = \frac{P(B/A_2)P(A_2)}{P(B)} = \frac{\frac{2}{5}\frac{1}{2}}{\frac{1}{2}} = 2/5$$

Ejemplo 2.- Supongamos en el ejemplo anterior de las baterías que el fabricante inspecciona una de las baterías y ésta es defectuosa y se quiere calcular a partir de dicho conocimiento la probabilidad de que la batería proceda de cada una de las tres plantas. En este caso

$$P(E_1/D) = \frac{P(D/E_1)P(E_1)}{\sum_{i=1}^{3} P(D/E_i)P(E_i)} = \frac{\frac{500}{4000}0.020}{0.02125} = 0.117647$$

$$P(E_2/D) = \frac{P(D/E_2)P(E_2)}{\sum_{i=1}^{3} P(D/E_i)P(E_i)} = \frac{\frac{2000}{4000}0.015}{0.02125} = 0.352941$$

$$P(E_2/D) = \frac{P(D/E_3)P(E_3)}{\sum_{i=1}^{3} P(D/E_i)P(E_i)} = \frac{\frac{1500}{4000}0.030}{0{,}02125} = 0.529412$$

A priori las probabilidades iniciales para cada planta eran 0.125, 0.5, 0.375, respectivamente, pero después del conocimiento de que la batería era defectuosa las probabilidades se han modificado a 0.117647, 0.352941 y 0.529412, respectivamente.

3. Independencia de sucesos

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $A \in \mathcal{A}$ con $P(A) > 0$. Como ya hemos comentado, la ocurrencia del suceso A puede alterar la probabilidad de ocurrencia de cualquier otro suceso $B \in \mathcal{A}$. Al estudiar dichas probabilidades, pueden darse los siguientes casos:

1. $P(B/A) \neq P(B)$, es decir la ocurrencia del suceso A modifica la probabilidad de ocurrencia de B. Diremos entonces que el suceso B **depende** del suceso A.

   Si $P(B/A) > P(B)$ se dice que el suceso A **favorece** al B.

   Si $P(B/A) < P(B)$ se dice que el suceso A **desfavorece** al B.

2. Si $P(B/A) = P(B)$, es decir, la ocurrencia del suceso A no tiene ningún efecto sobre el suceso A, se dice que el suceso B es **independiente** del suceso A.

Teorema: Caracterización de independencia

Sea $A \in \mathcal{A}$ con $P(A) > 0$. Un suceso B es independiente de $A \Longleftrightarrow P(A \cap B) = P(A) \cdot P(B)$

<u>Demostración</u>

$\Longrightarrow$) B independiente de $A \Rightarrow \; P(B/A) = \dfrac{P(A \cap B)}{P(A)} = P(B) \;\Rightarrow\; P(A \cap B) = P(A)P(B)$

$\Longrightarrow$) $P(A \cap B) = P(A)P(B)$ y $P(A) > 0 \Rightarrow P(B/A) = P(B) \;\Rightarrow\;$ B independiente de A.

Corolario

Este teorema pone de manifiesto la simetría de la definición, es decir, si $P(A) > 0$ y $P(B) > 0$, A es independiente de B si y sólo si B lo es de A y diremos, en general, que A y B son independientes.

<u>**Notas**</u>

- Un suceso nulo, $P(B) = 0$, es independiente de cualquier otro suceso, ya que si A es tal que $P(A) > 0$, se tiene

$$P(A \cap B) \leq P(B) = 0 = P(A)P(B) \Longrightarrow P(A \cap B) = 0$$

- Un suceso seguro, $P(B) = 1$, es independiente de cualquier otro ya que $P(B/A) = 1$.

Proposición.- Si A y B son independientes, entonces

1. A y $\overline{B}$ son independientes.
2. $\overline{A}$ y B son independientes.
3. $\overline{A}$ y $\overline{B}$ son independientes.

Demostración. Podemos suponer $P(A) \neq 0$ o 1 y $P(B) \neq 0$ o 1, ya que los complementarios en tal caso son también nulos o seguros y la independencia está garantizada.

1. $P(A \cap \overline{B}) = P(A) \cdot P(\overline{B}/A) = P(A)(1 - P(B/A)) = P(A)(1 - P(B)) = P(A) \cdot P(\overline{B})$
2. Análoga al anterior intercambiando los papeles de A y B. $P(\overline{A} \cap B) = P(B) \cdot P(\overline{A}/B) = P(B)(1 - P(A/B)) = P(B)(1 - P(A)) = P(B) \cdot P(\overline{A})$
3. $P(\overline{A} \cap \overline{B}) = P(\overline{A}) \cdot P(\overline{B}/\overline{A}) = P(\overline{A})(1 - P(B/\overline{A})) = P(\overline{A})(1 - P(B)) = P(\overline{A}) \cdot P(\overline{B})$
   Se puede hacer directamente de 1) dado que A y $\overline{B}$ son independientes.

La definición de independencia puede extenderse a una familia de sucesos y en esta extensión caben dos definiciones:

Definición 1: Independencia dos a dos.- Dado un espacio probabilístico $(\Omega, \mathcal{A}, P)$ y una clase de sucesos $\mathcal{U} \subset \mathcal{A}$ no vacía, diremos que sus sucesos son **independientes dos a dos**, si

$$\forall A, B \in \mathcal{U}, \;\; A \neq B, \quad A \text{ y } B \text{ son independientes}$$

Definición 2: Independencia mtua.- Dado un espacio probabilístico $(\Omega, \mathcal{A}, P)$ y una clase de sucesos $\mathcal{U} \subset \mathcal{A}$ no vacía, diremos que sus sucesos son **mutuamente (completamente o totalmente) independientes** o simplemente **independientes**, si para toda subcolección finita $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$ de suceso distintos de $\mathcal{U}$ se verifica

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$$

Nota.- Está claro que la independencia mutua implica la independencia dos a dos, pero el recíproco no es cierto en general como probamos a continuación.

Ejemplo.- Se lanzan dos dados y se consideran los sucesos:

A: salir impar en el primero
B: salir impar en el segundo
C: la suma de los resultados es impar

$P(A) = P(B) = P(C) = 1/2$

$A \cap B = \{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\}$
$P(A \cap B) = \frac{9}{36} = \frac{1}{4} = P(A)P(B)$

$A \cap C = \{(1,2),(1,4),(1,6),(3,2),(3,4),(3,6),(5,2),(5,4),(5,6)\}$
$P(A \cap C) = \frac{9}{36} = \frac{1}{4} = P(A)P(C)$

$B \cap C = \{(2,1),(4,1),(6,1),(2,3),(4,3),(6,3),(2,5),(4,5),(6,5)\}$
$P(B \cap C) = \frac{9}{36} = \frac{1}{4} = P(B)P(C)$

Sin embargo

$$P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$$

luego los sucesos son dos a dos independientes, pero no mutuamente independientes.