Probabilidad Condicionada

IES O Couto

curso 2019-2020

Antes de la definición axiomática debida a Kolmogorov, se consideró la probabilidad de un suceso como límite de la frecuencia relativa del suceso.
El problema que vamos a tratar en este tema, es el de formalizar la idea intuitiva de que la información aportada por el hecho de que ya haya ocurrido un suceso, ha de ser recogida cambiando el espacio de probabilidad de partida.
Veámoslo con un ejemplo:

Consideremos un experimento aleatorio consistente en lanzar un dado, y el espacio muestral $\Omega = \{1,2,3,4,5,6\}$. Sea B el suceso "ser par", es decir $B = \{2,4,6\}$:

- Suponiendo que todos los casos son equiprobables, por la ley de Laplace $\mathcal{P}(B) = \frac{3}{6} = \frac{1}{2}$
- Supongamos ahora que disponemos de la siguiente información: 'En el lanzamiento se obtuvo un número mayor que 3', es decir se verificó el suceso $A = \{4,5,6\}$.
  En esta situación, la probabilidad de B es ahora $\frac{2}{3}$.
  Esta probabilidad, en la que se utiliza el conocimiento de que se verificó A, se representa $\mathcal{P}(B/A)$

La noción de condicionalidad nace de la necesidad de querer utilizar la información que aporta saber que ha ocurrido A, para medir la probabilidad de la ocurrencia de B.
Así, supongamos que realizamos un experimento N veces. Si quisiéramos aproximar la probabilidad de un suceso B, usaríamos su frecuencia relativa $f_B = \frac{N_B}{N}$.
Ahora bien, si realizamos el experimento y tenemos información de que se verificó un suceso A, para calcular la probabilidad de ocurrencia de B usando dicha información, lo natural es recurrir al cociente $f_{B/A} = \frac{N_{A\cap B}}{N_A}$, que mide la frecuencia de realización de B con respecto la frecuencia de veces que se presenta el suceso A:

$$f_{B/A} = \frac{N_{A\cap B}}{N_A} = \frac{\frac{N_{A\cap B}}{N}}{\frac{N_A}{N}} = \frac{f_{A\cap B}}{f_A} \implies f_{B/A} = \frac{f_{A\cap B}}{f_A}$$

Definición

- Llamamos probabilidad del suceso B condicionada por A, y la denotamos por $\mathcal{P}(B/A)$, a la probabilidad de que ocurra B sabiendo que ha ocurrido A.
- Teniendo en cuenta la idea de probabilidad como límite de frecuencias relativas, tiene sentido definir la probabilidad del suceso B condicionada por A como:

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

- Está definida siempre que $\mathcal{P}(A) > 0$ (es decir, si $\mathcal{P}(A) = 0$ no se define).

Veamos que $\mathcal{P}(\cdot/A)$ es en efecto una medida de probabilidad comprobando que verifica los axiomas de Kolmogorov:

I) $p(B/A) \geq 0 \quad \forall B \subset \Omega$, pues $\mathcal{P}(A \cap B) \geq 0$ al ser $\mathcal{P}$ una probabilidad

II) $\mathcal{P}(\Omega/A) = 1$. En efecto

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

III) Sea $(B_n)_{n\in\mathbb{N}}$, con $B_i \cap B_j = \emptyset \quad \forall i \neq j$, entonces $\mathcal{P}(\bigcup_n B_n/A) = \sum_n \mathcal{P}(B_n/A)$, en efecto

$$\mathcal{P}(\bigcup_n B_n/A) = \frac{\mathcal{P}\Big((\bigcup_n (B_n \cap A)\Big)}{\mathcal{P}(A)} \underbrace{=}_{*} \frac{\sum_n \mathcal{P}(B_n \cap A)}{\mathcal{P}(A)} = \sum_n \mathcal{P}(B_n/A)$$

$$(*) \quad (B_i \cap A) \cap (B_j \cap A) = \emptyset \quad \forall i \neq j$$

$\mathcal{P}(\cdot/A)$ verifica todas las propiedades de una probabilidad:

- $\mathcal{P}(\emptyset/A) = 0$
- $\mathcal{P}(\overline{B}/A) = 1 - \mathcal{P}(B/A)$
- Si $B_1 \subset B_2$ entonces $\mathcal{P}(B_1/A) \leq \mathcal{P}(B_2/A)$
- $\mathcal{P}(B_1 \cup B_2/A) = \mathcal{P}(B_1/A) + \mathcal{P}(B_2/A) - \mathcal{P}(B_1 \cap B_2/A)$
- Si $B \subset A$ entonces $\mathcal{P}(B/A) = \frac{\mathcal{P}(B)}{\mathcal{P}(A)}$
- Si $A \subset B$ entonces $\mathcal{P}(B/A) = 1$
- Si $A \cap B = \emptyset$ entonces $\mathcal{P}(B/A) = 0$

Teorema de la Probabilidad Compuesta

Utilizando asimismo la definición de probabilidad condicionada puede obtenerse una forma de hallar la probabilidad de la intersección de dos sucesos. Nos la propociona el siguiente teorema.

Teorema

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad, y $A \in \mathcal{A}$ un suceso con $\mathcal{P}(A) > 0$, entonces:

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

La demostración es inmediata aplicando la definición de probabilidad condicionada.
El anterior teorema admite la siguiente generalización:

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad y $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ con $\mathcal{P}(\bigcap\limits_{i=1}^{n-1} A_i) > 0$, entonces

$$\mathcal{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathcal{P}(A_1) \cdot \mathcal{P}(A_2/A_1) \cdot \mathcal{P}(A_3/A_1 \cap A_2) \cdot \ldots \cdot \mathcal{P}\left(A_n/\bigcap_{i=1}^{n-1} A_i\right)$$

Definición

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad, y $A,\ B \in \mathcal{A}$.

- Los sucesos A y B son independientes si $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cdot \mathcal{P}(B)$.
- Los sucesos A y B son dependientes si $\mathcal{P}(A \cap B) \neq \mathcal{P}(A) \cdot \mathcal{P}(B)$.

Es importante observar que en la definición de independencia (o dependencia) de sucesos no se imponen restricciones sobre los valores de $\mathcal{P}(A)$ o $\mathcal{P}(B)$ (mientras que sí se hacía en la definición de probabilidad condicionada)

Propiedades

- Si $\mathcal{P}(A) > 0$ entonces A y B son independientes si y sólo si $\mathcal{P}(B/A) = \mathcal{P}(B)$
  Es importante destacar la hipótesis $\mathcal{P}(A) > 0$ pues si $\mathcal{P}(A) = 0$ no está definida la función $\mathcal{P}(\cdot/A)$. El caso $\mathcal{P}(A) = 0$ queda resuelto en la siguiente propiedad.
- Si $\mathcal{P}(A) = 0$ o $\mathcal{P}(A) = 1$ entonces A y B son sucesos independientes.
- Dados dos sucesos A y B, las siguientes afirmaciones son equivalentes:
  - A y B son independientes
  - A y $\overline{B}$ son independientes
  - $\overline{A}$ y B son independientes
  - $\overline{A}$ y $\overline{B}$ son independientes
- Si los sucesos A y B son incompatibles ($A \cap B = \emptyset$), con $\mathcal{P}(A) > 0$ y $\mathcal{P}(B) > 0$, entonces son dependientes.
  Por la incompatilidad si ocurre A necesariamente no puede ocurrir B, luego el hecho de que haya ocurrido A proporciona la información de la no ocurrencia de B. Es decir, la ocurrencia de B depende de la *no* ocurrencia de A.
  Recíprocamente, si A y B son independientes con $\mathcal{P}(A) > 0$ y $\mathcal{P}(B) > 0$, según el razonamiento anterior no pueden ser incompatibles.

Definición: Sucesos mutuamente independientes

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad, y $\{A_1, \ldots, A_n\}$ una colección de sucesos de $\mathcal{A}$. Se dice que los sucesos de la colección $\{A_1, \ldots, A_n\}$ son mutuamente independientes si para cualquier subconjunto $\{A_{i_1}, \ldots, A_{i_k}\}$ con $1 \leq k \leq n$, $i_l \neq i_m \; \forall l, m \in \{1, \ldots, k\}$ se tiene

$$\mathcal{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathcal{P}(A_{i_1}) \cdot \ldots \cdot \mathcal{P}(A_{i_k})$$

Los resultados relativos al teorema de la probabilidad compuesta son frecuentemente aplicados cuando se trabaja con varios experimentos aleatorios sucesivos, pues en ocasiones las condiciones iniciales de cada uno de ellos pueden venir alteradas por los resultados del experimento anterior. Si ello es así, se dice que los experimentos son dependientes, y en caso contrario se dirán independientes.
Supongamos que realizamos dos experimentos aleatorios con espacios de probabilidad asociados $(\Omega_1, \mathcal{A}_1, \mathcal{P}_1)$ y $(\Omega_2, \mathcal{A}_2, \mathcal{P}_2)$ respectivamente.

- Si los experimentos aleatorios son independientes (por ejemplo, cuando son físicamente independientes) sobre $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ podemos considerar la probabilidad:

$$\mathcal{P}(A_1 \times A_2) = \mathcal{P}_1(A_1) \cdot \mathcal{P}_2(A_2)$$

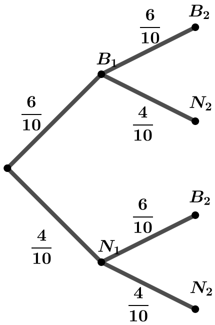
- Si los experimentos aleatorios son dependientes sobre $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$, se considerará la medida de probabilidad:

$$\mathcal{P}(A_1 \times A_2) = \mathcal{P}_1(A_1) \cdot \mathcal{P}_2(A_2/A_1)$$

Ejemplo 1

Se tiene una urna con seis bolas blancas y cuatro negras. Se extraen al azar dos bolas con reemplazamiento, y se desea construir el espacio de probabilidad asociado al experimento.

Podemos considerar este experimento como la composición de dos experimentos aleatorios consistentes en extraer una bola de la urna.
El espacio muestral es $\Omega = \{B_1 B_2,\ B_1 N_2,\ N_1 B_2,\ N_1 N_2\}$, y para asignar las probabilidades de cada uno de los sucesos elementales.
Como al haber reemplazamiento, la composición de la urna no varía, y en el diagrama en árbol esto se traduce en que las probabilidades de cada una de las ramas no dependen del origen de la ramificación

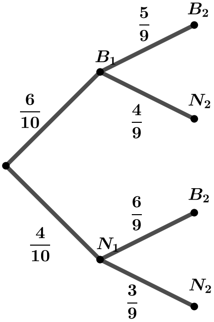


$$\mathcal{P}(B_1 \cap B_2) = \mathcal{P}(B_1) \cdot \mathcal{P}(B_2) = \frac{36}{100}$$
$$\mathcal{P}(B_1 \cap B_2) = \mathcal{P}(B_1) \cdot \mathcal{P}(B_2) = \frac{24}{100}$$
$$\mathcal{P}(B_1 \cap N_2) = \mathcal{P}(B_1) \cdot \mathcal{P}(N_2) = \frac{24}{100}$$
$$\mathcal{P}(N_1 \cap N_2) = \mathcal{P}(N_1) \cdot \mathcal{P}(N_2) = \frac{16}{100}$$

Ejemplo: Experimentos Independientes

Ejemplo 2

Si realizamos ahora el experimento anterior, pero sin reemplazamiento de las bolas, la composición de la urna cambia en la segunda extracción, con lo cual, las probabilidades de las ramas correspondientes a la segunda extracción no permanecen invariantes

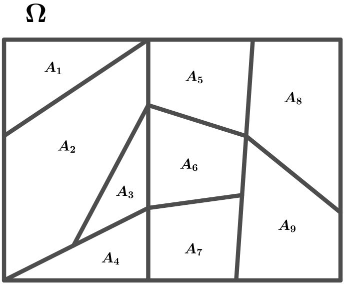


$$\mathcal{P}(B_1 \cap B_2) = \mathcal{P}(B_1) \cdot \mathcal{P}(B_2/B_1) = \frac{30}{90}$$
$$\mathcal{P}(B_1 \cap B_2) = \mathcal{P}(B_1) \cdot \mathcal{P}(B_2/B_1) = \frac{24}{90}$$
$$\mathcal{P}(B_1 \cap N_2) = \mathcal{P}(B_1) \cdot \mathcal{P}(N_2/B_1) = \frac{24}{90}$$
$$\mathcal{P}(N_1 \cap N_2) = \mathcal{P}(N_1) \cdot \mathcal{P}(N_2/N_1) = \frac{12}{90}$$

Ejemplo: Experimentos Dependientes

Definición

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio de probabilidad y $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ una familia numerable de sucesos. Se dice que constituye un sistema completo de sucesos si $A_i \cap A_j = \emptyset \quad \forall i \neq j$ y $\bigcup_{n\in\mathbb{N}} A_n = \Omega$.



Ejemplo: Sistema completo de sucesos

Teorema

Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Si tenemos un sistema completo de sucesos $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, con $\mathcal{P}(A_n) > 0 \quad \forall n \in \mathbb{N}$, entonces:

$$\mathcal{P}(B) = \sum_{n=1}^{+\infty} \mathcal{P}(B/A_n) \cdot \mathcal{P}(A_n) \quad \forall B \in \mathcal{A}$$

Teniendo en cuenta las siguientes igualdades de conjuntos

$$B = B\cap\Omega = B\cap(\bigcup_{n\in\mathbb{N}} A_n) = \bigcup_{n\in\mathbb{N}}(B\cap A_n) \quad \text{con } (B\cap A_i)\cap(B\cap A_j) = \emptyset \quad \forall i \neq j$$

se tiene la que la probabilidad del suceso B es $\mathcal{P}(B) = \sum_{n\in\mathbb{N}} \mathcal{P}(B\cap A_n)$

Y como además $\mathcal{P}(A_n) > 0 \, \forall n \in \mathbb{N}$, y $\mathcal{P}(B\cap A_n) = \mathcal{P}(A_n)\cdot\mathcal{P}(B/A_n) \, \forall n \in \mathbb{N}$, entonces:

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

Teorema

Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Si tenemos un sistema completo de sucesos $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, con $\mathcal{P}(A_n) > 0 \quad \forall n \in \mathbb{N}$, entonces:

$$\mathcal{P}(A_i/B) = \frac{\mathcal{P}(A_i)\cdot\mathcal{P}(B/A_i)}{\sum_{n=1}^{+\infty}\mathcal{P}(B/A_n)\cdot\mathcal{P}(A_n)} \quad \forall B \in \mathcal{A}/\mathcal{P}(B) > 0,\ \forall i \in \mathbb{N}$$

Si suponemos que $\mathcal{P}(B) > 0$, entonces $\forall i \in \mathbb{N}$ se tiene que

$$\mathcal{P}(A_i/B) = \frac{\mathcal{P}(A_i \cap B)}{\mathcal{P}(B)} = \frac{\mathcal{P}(B/A_i)\cdot\mathcal{P}(A_i)}{\mathcal{P}(B)}$$

Por definición de probabilidad condicionada: $\mathcal{P}(A_i \cap B) = \mathcal{P}(B/A_i)\cdot\mathcal{P}(A_i)$.
Sustituyendo el denominador por la fórmula de la Probabilidad Total tenemos

$$\mathcal{P}(A_i/B) = \frac{\mathcal{P}(A_i)\cdot\mathcal{P}(B/A_i)}{\sum_{n=1}^{+\infty}\mathcal{P}(B/A_n)\cdot\mathcal{P}(A_n)} \quad \forall B \in \mathcal{A}\ /\ \mathcal{P}(B) > 0,\ \forall i \in \mathbb{N}$$

Intepretación del Teorema de Bayes

Supongamos estamos estudiando algún fenómeno aleatorio, y que los sucesos A_n son considerados como hipótesis sobre alguna cuestión relacionada:

- Si tenemos cierta información reflejada por las probabilidades $\mathcal{P}(A_n) > 0$, $n \in \mathbb{N}$, y se produce el suceso B, entonces el teorema de Bayes nos informa de la probabilidad de que un A_n concreto haya sido la causa de la ocurrencia de B.
- Es decir, es un teorema que nos proporciona un método de medición de las probabilidades de las causas a partir de los efectos observados.

Por esta razón:

- A las probabilidades $\mathcal{P}(A_n)$ se las denomina probabilidades a priori.
- A las probabilidades $\mathcal{P}(A_n/B)$ se las denomina probabilidades a posteriori.
- A las $\mathcal{P}(B/A_n)$ se las denomina verosimilitudes.

Ejemplo 3

Supongamos una población de individuos en la que el 1 % tiene una enfermedad. Empleamos un test que permite detectar la patología en una persona enferma 8 de cada 10 (test positivo), y detectar la no existencia de la enfermedad en un individuo sano 9 de cada 10 veces (test negativo) ¿Cuál será la probabilidad de que un individuo esté realmente enfermo si dio positivo en el test?

Sean los sucesos $A = \{\text{sujeto enfermo}\}$ y $B = \{\text{test positivo}\}$. Por el teorema de Bayes

$$\mathcal{P}(A/B) = \frac{\mathcal{P}(B/A)\mathcal{P}(A)}{\mathcal{P}(B/A)\mathcal{P}(A) + \mathcal{P}(B/\overline{A})\mathcal{P}(\overline{A})} = \frac{\frac{8}{10} \cdot \frac{1}{100}}{\frac{8}{10} \cdot \frac{1}{100} + \frac{1}{10} \cdot \frac{99}{100}} = \frac{8}{107}$$

Así, $\mathcal{P}(A/B) = 0'07476$ es la probabilidad de que un sujeto esté realmente enfermo cuando el test dio positivo.