Probabilidad Condicional e Independencia

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

Probabilidad Condicional

Probabilidad Condicional (1)

En muchas situaciones tendremos que estimar la probabilidad de un evento pero disponemos de alguna información adicional sobre su resultado.

Por ejemplo supongamos que arrojamos un dado (equilibrado) y nos preguntamos ¿Qué probabilidad le asignaríamos a sacar un dos, si supiéramos de antemano que el resultado será un número par?.

Para formalizar esta pregunta consideramos en el espacio muestral

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

los eventos

$$A = \text{sale un 2} \; = \{2\}$$

$$B = \text{sale un número par} \; = \{2, 4, 6\}$$

Entonces vamos a definir la probabilidad condicional de que ocurra el evento A sabiendo que ocurre el evento B que notaremos $P(A/B)$.

Probabilidad Condicional (2)

Si estamos en una situación como la anterior donde la definición clásica de Laplace se aplica podemos pensarlo del siguiente modo: los resultados posibles de nuestro experimento son ahora sólo los elementos de B (es decir: hemos restringido nuestro espacio muestral a B), mientras que los casos favorables son ahora los elementos de $A \cap B$ luego

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

Si dividimos numerador y denominador por $\#(\Omega)$, tenemos:

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

Aunque hemos deducido esta fórmula de la definición clásica de Laplace, la misma tiene sentido en general siempre que $P(B) > 0$.

Probabilidad Condicional:Definición

Adoptamos pues la siguiente definición:

Definición

Sea $(\Omega, \mathcal{E}, P)$ *un espacio de probabilidad. La probabilidad condicional* $P(A/B)$ *de un evento* A *suponiendo que ocurre el evento* B *se define por:*

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

siempre que $P(B) > 0$.

Probabilidad Condicional: Otra manera de pensarla (1)

Otra manera de comprender esta definición es la siguiente: para definir la probabilidad condicional $P(A/B)$ queremos reasignar probabilidades a los eventos $A \in \mathcal{E}$ de modo que se cumplan tres condiciones:

1. La función $A \mapsto P(A/B)$ definida sobre $\mathcal{E}$ debe ser una probabilidad (o sea satisfacer los requisitos de nuestra definición axiomática).
2. $P(A \cap B/B) = P(A/B)$ (Esta fórmula dice que la probabilidad condicional de que ocurran los eventos A y B simultaneamente sabiendo que ocurre B debe ser igual a la probabilidad condicional de A sabiendo que ocurre B).
3. Si $A \subset B$ la probabilidad condicional $P(A/B)$ debe ser proporcional a la probabilidad de A de modo que

$$P(A/B) = kP(A) \text{ si } A \subset B$$

siendo k una constante de proporcionalidad fija.

Probabilidad Condicional: Otra manera de pensarla (2)

Entonces a partir de estas dos condiciones tenemos:

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

y como queremos que $P(A/B)$ sea una probabilidad debe ser $P(\Omega/A) = 1$, luego

$$1 = kP(\Omega \cap B) = kP(B)$$

con lo que:

$$k = \frac{1}{P(B)}$$

y vemos que la definición

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

es la única que satisface estas condiciones.

Fórmila de la probabilidad total

Proposición

Consideramos una partición del espacio muestral Ω en eventos disjuntos $B_1, B_2, \ldots B_n, \ldots$ (finita o infinita numerable)

$$\Omega = \bigcup_{k \in \mathbb{N}} B_k, \quad B_i \cap B_j = \emptyset \text{ si } i \neq j$$

con $P(B_k) > 0$ para todo k, entonces

$$P(A) = \sum_{k=1}^{\infty} P(A/B_k) \cdot P(B_k)$$

Demostración

Por la σ-adivitividad y la definición de probabilidad condicional

$$P(A) = \sum_{k=1}^{\infty} P(A \cap B_k) = \sum_{k=1}^{\infty} P(A/B_k) \cdot P(B_k)$$

Independencia (1)

Definición

Decimos que el evento A es independiente del evento B con $P(B) > 0$ si

$$P(A/B) = P(A)$$

Intuitivamente este concepto significa que saber si el evento B ocurre o no, no nos dará una mejor estimación de la probabilidad de que ocurre el evento B que si no lo supiéramos.
Teniendo en cuenta la definición de la probabilidad condicional, vemos que la condición para que el evento A sea independiente de B es que:

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

Independencia (2)

Esta manera de escribir la definición tiene dos ventajas: se ve que tiene sentido aún si $P(B) = 0$, y muestra que los roles de los eventos A y B son simétricos. Reescribimos pues la definición en la siguiente forma:

Definición

Decimos que los eventos A y B son (estocásticamente) independientes si

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

OJO: ¡no confundir eventos independientes con eventos disjuntos!.

Un ejemplo sencillo

Consideramos el experimento de extraer una carta de un mazo de 48 cartas españolas, y consideramos los eventos:

- $A =$ "sale un 1".
- $B =$ 'sale una carta de espadas"

$$P(A) = \frac{4}{48} = \frac{1}{12}$$

$$P(B) = \frac{12}{48} = \frac{1}{4}$$

Entonces $A \cap B$ es "sale el uno de espadas" y

$$P(A \cap B) = \frac{1}{48} = \frac{1}{12} \cdot \frac{1}{4} = P(A) \cdot P(B)$$

Luego A y B son independientes.

Una aplicación a la ecología (1)

Supongamos que tenemos una población de animales en un territorio y queremos estimar cuántos animales hay. Un método posible es el de captura / recaptura. Se utiliza mucho para poblaciones de micro mamíferos y reptiles. Mediante trampas se capturan individuos que son marcados y devueltos a su ambiente. Después de un cierto período de tiempo, suficiente para que los marcados se mezclen con el resto de la población, se realiza una nueva captura

Nuestro espacio muestral Ω serán los individuos de la población. Consideramos, para un individuo elegido al azar, los eventos:
$A =$ "el animal es capturado en la primera captura (y marcado)."
$B =$ "el animal es capturado en la segunda captura."
$C = A \cap B =$ "el animal es capturado en la segunda captura y estaba marcado."

Llamemos n_A al número de elementos capturados en la primera captura, n_B a los capturados en la segunda captura, n_C a los capturados en la segunda captura y n_Ω al número total de individuos en la población. n_A, n_B y n_C son conocidos. Queremos determinar n_Ω.

Una aplicación a la ecología (2)

Si la población es grande, la probabilidad de un individuo de ser capturado será parecida a la frecuencia observada. Entonces podemos estimar las probabilidades

$$P(A) \approx \frac{n_A}{n_\Omega}, P(B) \approx \frac{n_B}{n_\Omega}, P(C) \approx \frac{n_C}{n_\Omega}$$

Ahora como las capturas son independientes esperamos que

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

por lo que tenemos la igualdad aproximada:

$$\frac{n_C}{n_\Omega} \approx \frac{n_A}{n_\Omega} \cdot \frac{n_B}{n_\Omega}$$

de donde podemos estimar el tamaño de la población como

$$n_\Omega \approx \frac{n_A \cdot n_B}{n_C}$$

Propiedades de la independencia de eventos

Proposición

Si A y B son eventos independientes, A y B^c también lo son.

Demostración.

Notamos que

$$A \cap B^c = A - B = A - A \cap B$$

Luego como $A \cap B \subset A$, y la hipótesis

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

Entonces

$$P(A \cap B^c) = P(A) \cdot [1 - P(B)] = P(A) \cdot P(B^c)$$

$\square$

Independencia con 3 eventos

Si tenemos 3 eventos A,B y C ¿cuándo diremos que son independientes?. No sólo vamos a querer que tengamos independencia de a pares:

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

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

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

Si no que también queremos por ejemplo que

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

Esto significa que queremos que A sea independiente de $B \cap C$ por lo que vamos a pedir que

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

o sea

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

Un ejemplo para entender esta definición

En $\Omega = \{1, 2, 3, 4, 5, \ldots, 36\}$ consideramos los eventos

$$A = \{1, 2, 3, 4, 5, 10, 11, 12, 13, 24, 25, 26, 27, 32, 33, 34, 35, 36\}$$

$$B = \{6, 7, 8, 9, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 28, 29, 30, 31\}$$

$$C = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18\}$$

De modo que:

$$A \cap B = \emptyset, A \cap B\{1, 2, 3, 4, 5, 10, 11, 12, 13\}, B \cap C = \{6, 7, 8, 9, 14, 15, 16, 17, 18\}$$

$$\#(A) = \#(B) = \#(C) = 18, \#(A \cap B) = \#(A \cap C) = \#(B \cap C) = 9$$

Luego

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

Ejemplo (2)

$$P(A \cap B) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = P(A) \cap P(B)$$

A y B no son indepedendientes.

$$P(A \cap C) = \frac{1}{4} = \frac{1}{4} \cdot \frac{1}{4} = P(A) \cap P(C)$$

A y C son independientes.

$$P(B \cap C) = \frac{1}{4} = \frac{1}{4} \cdot \frac{1}{4} = P(B) \cap P(C)$$

B y C son independientes.

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

pero

$$P(A) \cdot P(B) \cdot P(C) = 1/8$$

Generalización a familias arbitrarias de eventos

En general, la definición de independencia es la siguiente:

Definición

Decimos que una familia cualquiera de eventos $\mathcal{A} \subset \mathcal{E}$ *es independiente si*

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_n})$$

para cualquier elección de una cantidad finita A_{i_1}*, ...,* A_{i_n} *de eventos distintos de la familia* $\mathcal{A}$*.*

Una generalización de la propiedad que vimos antes

Ejercicio 9 de la práctica 2

Probar que si $A_1, \ldots, A_n$ son eventos independientes y $B_1, \ldots, B_n$ son tales que para cada $i = 1, \ldots, n$ se tiene $B_i = A_i$ o $B_i = A_i^c$ entonces los eventos $B_1, \ldots, B_n$ también resultan independientes.

En este ejercicio, $\mathcal{A} = \{A_1, \ldots, A_n\}$.

Parte II

El esquema de ensayos de Bernoulli

Ensayos de Bernoulli

Consideramos un experimento aleatorio con dos resultados que convencionalmente se llaman

- éxito (1)
- fracaso(0)

Lo podemos modelizar con el espacio de probabilidad $\Omega = \{0, 1\}$ donde

$$P(\{1\}) = p, P(\{0\}) = q = 1 - p$$

Por ejemplo: La clase pasada consideramos el caso de arrojar una moneda donde

$$p = q = \frac{1}{2}$$

Repitiendo el experimento

Ahora supongamos que repetimos el experimento n veces. Como vimos la vez pasada, el producto cartesiano

$$\Omega^n = \underbrace{\Omega \times \Omega \times \cdots \times \Omega}_{n \text{ veces}} = \{\omega = (\omega_1, \omega_2, \ldots, \omega_n) : \omega_i \in \Omega \text{ para todo } i\}$$

se puede usar como espacio muestral para modelar la repetición de este experimento n veces. Aquí ω_i representa el resultado del experimento en la i-ésima realización del experimento.

Notemos X_i al resultado de la i-ésima realización del experimento.

$$X_i(\omega) = \omega_i$$

$X_i : \Omega \to \mathbb{R}$ es otro ejemplo de una variable aleatoria con distribución

$$P\{X_i = 1\} = p, P\{X_i = 0\} = q = 1 - p$$

Las X_i se llaman variables de Bernoulli. Notación $X_i \sim \text{Be}(p)$.

Como asignamos probabilidades

¿Cómo pdemos asignar probabilidades en Ω^n? Si consideramos que las n realizaciones del experimento son independientes (una realización no afecta a la otra) tenemos

$$
\begin{aligned}
P\{(\omega_1, \omega_2, \ldots, \omega_n)\} &= P\{X_1 = \omega_1, X_2 = \omega_2, \ldots, X_n = \omega_n\} \\
&= P\{X_1 = \omega_1\} \cdot P\{X_2 = \omega_2\} \cdots P\{X_n = \omega_n\} \\
&= p^k q^{n-k}
\end{aligned}
$$

donde k es el número de éxitos en los n-ensayos. Podemos escribir $k = S_n(\omega)$.

$$S_n = X_1 + X_2 + \ldots + X_n$$

es otro ejemplo de una variable aleatoria.

La suposición de independencia estuvo implícita en el ejemplo de tirar n veces la moneda de la clase anterior.

La distribución de Bernoulli

¿Cuál es la distribución de la variable S_n? Para que $S_n = k$ necesitamos k éxitos y $n-k$ fracasos entre los n ensayos. Luego hay exactamente $\binom{n}{k}$ puntos de Ω^n donde esto ocurre, y todos ellos tienen probabilidad $p^k q^{n-k}$. Por lo tanto

$$P\{S_n = k\} = \binom{n}{k} p^k q^{n-k} \quad \text{donde } q = 1 - p$$

Se conoce como distribución Binomial o distribución de Bernoulli de parámetro p.

Notación: $S_n \sim \mathrm{Bi}(n, p)$.

La clase pasada vimos el caso particular $p = 1/2$.

Un ejemplo

Supongamos que al producir una lamparita hay un 1 % de fallos. Se producen 100 lámparitas. ¿Cuál es la probabilidad de tener exactamente 3 falladas?
Ponemos $p = 0{,}01 \Rightarrow q = 0{,}99$ (¡notemos que en este ejemplo el "éxito" es que la lamparita esté fallada!). Queremos calcular $P = \{S_n = 3\}$.

$$P\{S_n = 20\} = \binom{100}{3} \cdot 0{,}01^3 \cdot 0{,}99^{97} \approx 0{,}06099165807528455$$

Programita en Python, usando SciPy

```
from scipy.stats import binom
k=3
n=100
p=0.01
print(binom.pmf(k, n, p))
```

binom.pmf = "probability mass function" de la distribución binomial.

Repitiendo infinitas veces el experimento

Como la clase pasada, podemos pensar también en repetir infinitas veces el experimento. El espacio muestral sería

$$\Omega^{\infty} = \{\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) : \omega_i \in \Omega \text{ para todo } i \in \mathbb{N}\}$$

Como este espacio es infinito no numerable, van a aparecer las mismas dificultades técnicas que encontramos en el caso $p = 1/2$. Para resolverlas se necesitan herramientas de la teoría de la medida (no profundizaremos en esto por el momento).

Parte III

Cadenas de Markov

Cadenas de Markov (homogéneas)

Consideramos un sistema que puede tener una cantidad finita de estados $\Omega = \{E_1, E_2, \ldots, E_n\}$ y que evoluciona con tiempo discreto $t \in \mathbb{N}_0$. Llamemos X_t al estado del sistema en el tiempo t.

La evolución del sistema estaría descripta por el espacio muestral

$$\Omega^\infty = \{X = (X_0, X_1, X_2, \ldots, X_t, \ldots) : X_t \in \Omega \text{ para todo } t \in \mathbb{N}_0\}$$

Suponemos que tenemos una cierta probabilidad de pasar del estado E_i al E_j

$$p_{ij} = P\{X_{t+1} = E_i / X_t = E_j\}$$

y que esta probabilidad es independiente de t (no varía en el tiempo)- Los números $P = (p_{ij})$ forman una matriz $den \times n$ que se denomina matriz de transición. Notamos que $0 \leq p_{ij} \leq 1$ y

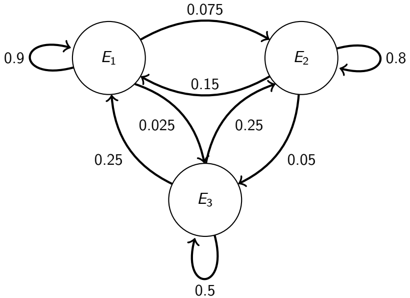
$$\sum_{i=1}^{n} p_{ij} = 1$$

(Las columnas de la matriz son vectores de probabilidad). P se dice una matriz estocástica.

Un ejemplo (sacado de Wikipedia)

Consideramos un modelo simple de 3 estados para un mercado financiero con 3 estados:

- E_1: Mercado en crecimiento.
- E_2: Mercado en decrecimiento.
- E_3: Mercado estancado.



$$P = \begin{bmatrix} 0{,}9 & 0{,}15 & 0{,}25 \\ 0{,}075 & 0{,}8 & 0{,}25 \\ 0{,}025 & 0{,}05 & 0{,}5 \end{bmatrix}$$

(Esto podría aplicarse a todo un mercado o a un activo particular negociado en ese mercado)

Application of Markov Process for Prediction of Stock Market Performance. Lakshmi G, Jyothi Manoj. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-6, March 2020.

Propiedades de la matriz de transición

Notamos que por la fórmula de probabilidad total:

$$P(X_{t+1} = E_i) = \sum_{j=1}^{n} P(X_{t+1} = E_i / X_t = E_j) \cdot P(X_t = E_j)$$

Esto quiere decir que si consideramos el vector de probabilidades

$$U_t = \begin{bmatrix} P(X_t = E_1) \\ P(X_t = E_2) \\ \dots \\ P(X_t = E_n) \end{bmatrix}$$

tendremos que

$$U_{t+1} = P \cdot U_t \quad \text{para todo } t \in \mathbb{N}$$

Entonces por inducción

$$U_t = P^t U_0$$

Comportamiento a largo plazo

En este ejemplo (diagonalizando la matriz), se puede ver que

$$\lim_{t\to+\infty} P^t = \begin{pmatrix} 0{,}625 & 0{,}625 & 0{,}625 \\ 0{,}3125 & 0{,}3125 & 0{,}3125 \\ 0{,}0625 & 0{,}0625 & 0{,}0625 \end{pmatrix}$$

Entonces la ecuación

$$U_t = P^t U_0$$

nos muestra que no importa cuál sea el estado inicial U_0, el sistema converge al estado estacionario:

$$U_\infty = \begin{pmatrix} 0{,}625 \\ 0{,}3125 \\ 0{,}0625 \end{pmatrix}$$

Esto significaría que el mercado va a estar un $62,5\,\%$ del tiempo en estado alcista, un $32,15\,\%$ en estado bajista y un $6,25\,\%$ estancado. Este vector es un autovector de autovalor 1 de P es decir:

$$PU_\infty = U_\infty = 1 \cdot U_\infty$$

Otros ejemplos de cadenas de Markov

- Juegos de tablero con dados, como el Monopoly.

monopoly.jpg

Monopoly in the view of mathematics
`http://www.bewersdorff-online.de/amonopoly/`

- El experimento de tirar infinitas veces la moneda es una cadena de Markov con dos estados $\{\text{cara}, \text{ceca}\}$. En este caso la matriz de transición es:

$$P = \begin{bmatrix} 0{,}5 & 0{,}5 \\ 0{,}5 & 0{,}5 \end{bmatrix}$$

  Esta matriz verifica $P^2 = P$. Luego $P^t = P$ para todo t, y tenemos que el estado estacionario es

$$U_\infty = \begin{bmatrix} 0{,}5 \\ 0{,}5 \end{bmatrix}$$