Repaso de probabilidad y estadística I

Alan Reyes-Figueroa

Introducción a la Ciencia de Datos (Aula 02) 14.enero.2021

Probabilidades

Construcción. Punto de partida: un experimento

- Resultado del experimento es $\omega \in \Omega$ $\rightsquigarrow$ *espacio muestral*.
- Interés en ciertos eventos A $\rightsquigarrow$ $\sigma-$*álgebra*
- Una probabilidad $\mathbb{P}$ es una función sobre ciertos eventos $\mathbb{P} : A \mapsto \mathbb{R}$.

Ejemplo 1

Experimento: lanzar un dado.

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

Algunos eventos

Representación	Evento
$A_1 = \{2, 4, 6\}$	obtener un número par
$A_2 = \{3\}$	obtener 3
$A_3 = \{1, 2, 4, 5\}$	obtener un número no múltiplo de 3

Ejemplo 2
Experimento: lanzar dos dados.
$\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),\ldots,(5,6),(6,6)\}$
Probablemente aquí sea más simple representarlo como
$\Omega = \{(a,b) : a,b \in [1..6]\} = [1..6] \times [1..6]$

Algunos eventos

Representación	Evento
$A_1 = \{(1,6),(2,5),(3,4),\ldots,(6,1)\}$	que los dados sumen 7
$A_2 = \{(1,3),(3,1),\ldots,(6,3),(3,6)\}$	que aparezca al menos un 3

Otros espacios asociados: $\Omega_1 = [1..6]$, ¿Cuál es el mínimo de los dos dados?

Otros ejemplos (para pensar)

Especificar un espacio muestral para los siguientes experimentos:

a) Lanzar una moneda.

b) Lanzar una moneda hasta que aparezca "cruz".

c) Distancia recorrida por un automóvil con un litro de gasolina.

d) Señal de radio que se recibe durante dos segundos.

e) Juego entre tres jugadores: P, Q y R. El juego consiste en jugar partidas por parejas, comenzando P contra Q. Quien gane un partida juega con el otro jugador, hasta que uno de los jugadores gane dos partidas consecutivas, ganando entonces el juego.

Pregunta: ¿Cómo definir $\mathbb{P}$? ¿Cómo interpretarla?

Definición (Espacio de probabilidad)

*Un **espacio de probabilidad** es una estructura* $(\Omega, \mathcal{F}, \mathbb{P})$*, donde*

- Ω *es un conjunto (no vacío). Los elementos* $\omega \in \Omega$ *se llaman eventos.*
- $\mathcal{F} \subseteq \Omega$ *es una* $\sigma-$*álgebra.*
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ *es una medida de probabilidad.*

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Definición

*Una $\sigma-$**álgebra** $\mathcal{F}$ sobre un conjunto Ω es una colección de subconjuntos de Ω que satisface:*

- $\Omega \in \mathcal{F}$;
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ *(es cerrada bajo complementos);*
- $A_i \in \mathcal{F}$, *para* $i = 1, 2, \ldots \Rightarrow \bigcup_i A_i \in \mathcal{F}$ *(es cerrada bajo uniones enum).*

Definición

*Una función $\mathbb{P} : \mathcal{F} \to [0, 1]$ es una **medida de probabilidad** si*

- $\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\Omega) = 1$;
- *para cualquier colección enumerable de eventos exclusivos $E_i \in \mathcal{F}$, vale*
$$\mathbb{P}\Big(\bigcup E_i\Big) = \sum \mathbb{P}(E_i) \ \ \textit{(enumerablemente aditiva)}.$$

Axiomas

Axiomas de la probabilidad, introducidos por Kolmogorov en 1933.

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de medida con $\mathbb{P}(E)$ la probabilidad de un evento $E \in \mathcal{F}$. Asumimos los siguientes supuestos para $\mathbb{P}$:

Axiomas

1. $\mathbb{P}(E) \geq 0, \forall E \in \mathcal{F}$ *(no-negativa).*
2. $\mathbb{P}(E)$ *es siempre finita, y* $\mathbb{P}(\Omega) = 1$ *(unitariedad).*
3. *Cualquier colección enumerable y mutuamente excluyente de eventos* $E_i \in \mathcal{F}$*, satisface*
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i), \qquad (\sigma\text{-aditiva}).$$

Propiedades

Si $\mathbb{P}$ es una medida de probabilidad sobre Ω, entonces

1. *(Monotonicidad) Si $A \subseteq B$ son eventos, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

2. *(Conjunto vacío) $\mathbb{P}(\emptyset) = 0$.*

3. *(Complemento) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, para todo evento $A \in \mathcal{F}$.*

4. *(Cotas para $\mathbb{P}$) Para todo evento $E \in \mathcal{F}$, $0 \leq \mathbb{P}(E) \leq 1$.*

5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Consecuencias

1. (Monotonicidad) Si $A \subseteq B$ son eventos en $\mathcal{F}$, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.

   Prueba:
   Definamos $E_1 = A$, $E_2 = B - A$, y $E_i =$ para $i = 3, 4, \ldots$. Entonces, por σ-aditividad (axioma 3),

$$\mathbb{P}(A) + \mathbb{P}(B - A) + \sum_{i \geq 3} \mathbb{P}(E_i) = P(B).$$

   Como el lado izquierdo anterior es una suma de términos no-negativos (axioma 1), entonces

$$\mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(B - A) + \sum_{i \geq 3} \mathbb{P}(E_i) = \mathbb{P}(B).$$

Consecuencias

1. (Monotonicidad) Si $A \subseteq B$ son eventos en $\mathcal{F}$, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.

   Prueba:
   Definamos $E_1 = A$, $E_2 = B - A$, y $E_i =$ para $i = 3, 4, \ldots$. Entonces, por σ-aditividad (axioma 3),

$$\mathbb{P}(A) + \mathbb{P}(B - A) + \sum_{i \geq 3} \mathbb{P}(E_i) = \mathbb{P}(B).$$

   Como el lado izquierdo anterior es una suma de términos no-negativos (axioma 1), entonces

$$\mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(B - A) + \sum_{i \geq 3} \mathbb{P}(E_i) = \mathbb{P}(B).$$

2. (Complemento) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, para todo evento $A \in \mathcal{F}$.

   Prueba:
   A y $A^c = \Omega - A$ forman una partición de Ω. Por σ-aditividad (axioma 3) y el axioma 2

   $$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1,$$

   luego $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

3. $\mathbb{P}(\emptyset) = 0$.

   Prueba:
   $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^c) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.

Consecuencias

4. $0 \leq \mathbb{P}(E) \leq 1$, para todo evento E.

   Prueba:

   $\mathbb{P}(E) \geq 0$ por el axioma 1. Además, $E \subseteq \Omega$, y la monotonicidad de $\mathbb{P}$ implican $\mathbb{P}(E) \leq \mathbb{P}(\Omega) = 1$.

5. (Principio de Inclusión-Exclusión) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

   Prueba:
   Observe que $\mathbb{P}(A - B) + \mathbb{P}(A \cap B) = \mathbb{P}(A)$ (por aditividad). Luego $\mathbb{P}(A - B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$. Similarmente, $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Ahora, $A \cup B$ es la unión disjunta de $A - B$, $B - A$ y $A \cap B$. Por aditividad,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A - B) + \mathbb{P}(B - A) + P(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Caso finito

Sea $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$.
Distribución de conteo o distribución uniforme: Corresponde a elegir un elemento al azar.
Para cada $A \subseteq \Omega$, se tiene

$$\mathbb{P}(A) = |A|/|\Omega| = |A|/k.$$

En particular, sin $A_i = \{\omega_i\}$, entonces

$$\mathbb{P}(\omega_i) = \mathbb{P}(\{\omega_i\}) = 1/k.$$

Caso general: Suponga que $\mathbb{P}(\omega_i) = \mathbb{P}(\{\omega_i\}) = p_i$, para $i = 1, 2, \ldots, k$.
Entonces

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i$$

Ejemplo

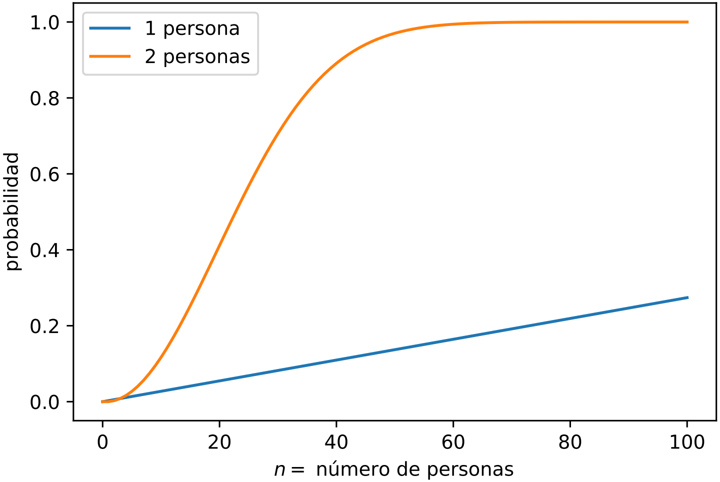
1. Calcula la probabilidad que en un grupo de n personas hay al menos una que cumple años el 14 de enero.

2. Calcula la probabilidad que en un grupo de n personas hay al menos dos personas que cumplen en el mismo día.

Ejemplo

n	probilidad 1 persona 14/enero	probabilidad 2 personas mismo día
0	0	0
1	0.002739	0
5	0.013698	0.027135
10	0.027397	0.116948
20	0.054794	0.411438
30	0.082191	0.706316
40	0.109589	0.891231
50	0.136986	0.970373
60	0.164383	0.994122
70	0.191780	0.999159

Ejemplo



1 persona
2 personas
probabilidad
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
n = número de personas

Caso $\Omega \subseteq \mathbb{R}^d$

Distribución uniforme:
Experimento: Elegir un número al azar de $[0, 2]$.
Tenemos $\Omega = [0, 2]$.
$A = [0, 1]$ $\qquad \mathbb{P}(A) = 1/2$.
$B = [0.4, 1]$ $\qquad \mathbb{P}(B) = 0.6/2 = 0.3$.

En general, para $A \subseteq \Omega$

$$\mathbb{P}(A) = \frac{\int_A dx}{\int_\Omega dx}.$$

¿Se puede calcular $\mathbb{P}$ siempre? No.

- Se requiere que $\int_\Omega dx < \infty$.
- Tenemos que limitarnos a conjuntos donde $\int_A dx$ existe.

Ejemplos

1. Se elige al azar un punto en un cuadrado con lado 4 cm. Calcula la probabilidad de que esté a una distancia menor de uno cm. de alguna de las esquinas.

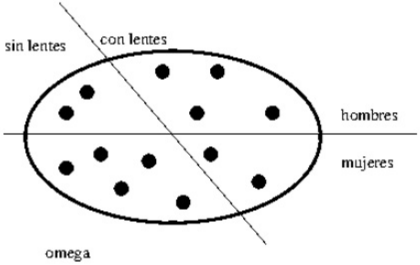
2. Dos estudiantes quieren ir a comer juntos. Se citan entre las 7 y las 8 de la noche y están dispuestos a esperar a lo más 10 minutos. ¿Cuál es la probabilidad de que puedan ir a comer si sus horas de llegada son uniformes entre las 7 y las 8?

Interpretación de $\mathbb{P}$

- A partir del experimento elegir algo al azar.
- Probabilidades como límite de frecuencias relativas de ocurrencia (**enfoque frequentista**)
- Por medio de apuestas: probabilidades como creencias (base del **enfoque bayesiano**)
- Sistema axiomático (Kolmogorov, 1933).

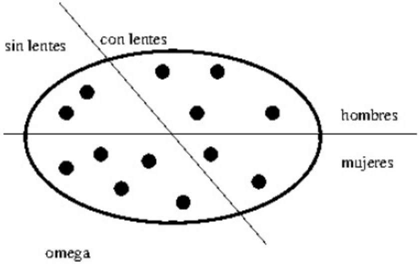
En áreas como computación e inteligencia artificial, se han elaborado otros sistemas axiomáticos (fuzzy sets, Dempster-Shaffer, $\ldots$)

Conceptos derivados: Probabilidad condicional



Se elige una persona al azar.
¿Cuál es la probabilidad que sea una persona con lentes? $\frac{6}{13}$.

Conceptos derivados: Probabilidad condicional



Se elige una persona al azar.
¿Cuál es la probabilidad que sea una persona con lentes? $\frac{6}{13}$.

Alguien dice que es un hombre: ¿cuál es ahora la probabilidad que sea una persona con lentes? $\frac{2}{3}$.

Definición

Si $\mathbb{P}(B) > 0$, entonces la probabilidad condicional de A dado B se define como

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

Probabilidad condicional

Observaciones:

- $\mathbb{P}(\cdot|B)$ define una nueva función de probabilidad sobre el espacio $\Omega' = B$.

- En consecuencia, $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$.

- Observar que no hay ninguna relación directa entre $\mathbb{P}(A|B)$ y $\mathbb{P}(A|B^c)$.

- Siempre podemos escribir $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$.
  (Esto no requiere el supuesto que $\mathbb{P}(B) > 0$) ¿Por qué?

Ejemplo

Experimento: Elegir al azar dos letras consecutivas de alguna palabra con alfabeto $T = \{a, b, c, d, e\}$.

Suponemos la siguiente distribución:

	a	b	c	d	e
a	0.10	0.05	0.10	0.04	0
b	0.01	0.01	0.10	0.01	0.04
c	0.02	0.05	0.05	0.10	0.01
d	0.04	0.10	0.01	0.01	0.02
e	0	0.10	0	0.01	0.02

¿Cuál es la probabilidad que la segunda letra seleccionada sea la “b” dado que sabemos que la anterior fue una vocal?

Ejemplo

Solución: Queremos calcular $\mathbb{P}(A|B)$, donde $B = \{\textit{primera letra es vocal}\}$ y $A = \{\ \textit{letra es b}\}$.

Entonces, de la definición de probabilidad condicional, tenemos

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

Pero, $\mathbb{P}(A \cap B) = \mathbb{P}(\{ab, eb\}) = \mathbb{P}(ab) + \mathbb{P}(eb) = 0.05 + 0.10 = 0.15$, y $\mathbb{P}(B) = \mathbb{P}(\{ab, bb, cb, db, eb\}) = 0.05 + 0.01 + 0.05 + 0.10 + 0.10 = 0.31$.

De allí que

$$P(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.15}{0.31} = 0.48387$$

Ley de la probabilidad total

Teorema (Ley de la probabilidad total, caso finito)

Dada una partición $\{B_i\}_{i=1}^n$ de Ω, tal que $\mathbb{P}(B_i) > 0$, $\forall i$, entonces

$$P(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

<u>Prueba</u>: $\Omega = \bigcup B_i$, ya que es una partición. Luego

$$A = A \cap \Omega = A \cap \bigcup B_i = \bigcup (A \cap B_i),$$

y los $A \cap B_i$ forman una partición de A. Por el axioma de aditividad, y la definición de probabilidad condicional

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

Ejemplo

Se tienen tres cajas, cada una conteniendo 100 cartas:
La caja 1 contiene 75 cartas rojas, y 25 cartas azules,
la caja 2 contiene 60 cartas rojas, y 40 cartas azules,
la caja 3 contiene 45 cartas rojas, y 55 cartas azules.

Se elige una de las cajas al azar, y luego se elige una carta dentro de la caja seleccionada.
¿Cuál es la probabilidad de elegir una carta roja?

Ejemplo

<u>Solución</u>: Considere los eventos $A =$ elegir carta roja, y

$$E_1 = \text{elegir caja } 1, \; E_2 = \text{elegir caja } 2, \; E_3 = \text{elegir caja } 3.$$

Sabemos que $\mathbb{P}(A|E_1) = \frac{75}{100}$, $\mathbb{P}(A|E_2) = \frac{60}{100}$ y $\mathbb{P}(A|E_3) = \frac{45}{100}$.

Entonces, por la ley de probabilidad total

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A|E_1)\,\mathbb{P}(E_1) + \mathbb{P}(A|E_2)\,\mathbb{P}(E_2) + \mathbb{P}(A|E_3)\,\mathbb{P}(E_3) \\
&= \frac{75}{100} \cdot \frac{1}{3} + \frac{60}{100} \cdot \frac{1}{3} + \frac{45}{100} \cdot \frac{1}{3} \\
&= \frac{245}{300} = 0.81666
\end{aligned}$$

Teorema (Regla de Bayes)

Si $\mathbb{P}(A), \mathbb{P}(B) > 0$*, entonces*

$$P(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Prueba: Por hipótesis, $\mathbb{P}(A), \mathbb{P}(B) > 0$, entonces

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{y} \quad \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Despejando $\mathbb{P}(A \cap B)$ de la segunda ecuación, $\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$, y sustituyendo en la primera

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

Ejemplo

Una companía ha desarrollado una prueba para detectar la presencia de SARS-CoV-2 (Covid-19). Se pretende que
$\mathbb{P}(\text{prueba es positiva}|\text{tiene covid}) = 0.97$ y
$\mathbb{P}(\text{prueba es negativa}|\text{no tiene covid}) = 0.98$.

Si el 1% de la población tiene Covid, calcular la probabilidad de decir que el test dé negativo cuando alguien sí tiene covid.

Ejemplo

Solución:

x: Test y: Real	No $= 0$	Sí $= 1$
No $= 0$		
Sí $= 1$		

Datos: $\mathbb{P}(x = 1) = 1/100$, $\mathbb{P}(y = 1|x = 1) = 0.97$, $P(y = 0|x = 0) = 0.98$. Queremos calcular $\mathbb{P}(x = 0|y = 1)$.

$$\begin{aligned}
\mathbb{P}(x = 0|y = 1) &= \frac{\mathbb{P}(y = 1|x = 0)\,\mathbb{P}(x = 0)}{\mathbb{P}(y = 1)} \\
&= \frac{\mathbb{P}(y = 1|x = 0)\,\mathbb{P}(x = 0)}{\mathbb{P}(y = 1|x = 0)\,\mathbb{P}(x = 0) + P(y = 1|x = 1)\,\mathbb{P}(x = 1)} \\
&= \frac{0.02(0.01)}{0.02(0.99) + 0.97(0.01)} = 0.006779
\end{aligned}$$

Conceptos derivados: Independencia

La idea de **independencia** es determinar si hay o no relación entre dos eventos A y B.
En otras palabra, si al conocer A, cambia nuestro conocimiento sobre B (o al conocer B cambia nuestro conocimiento sobre A).

¿Cómo determinar esta relación? Comparar $\mathbb{P}(A|B)$ con $\mathbb{P}(A)$.

Definición

*Si $\mathbb{P}(B) > 0$, decimos que A y B son **independientes** si $\mathbb{P}(A|B) = \mathbb{P}(A)$.*

Definición

*Dos eventos A y B son **independientes** si, y sólo si,*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Referencias

- Kai-Lai Chung. *A Course in Probability*.