

Contingencia e información en la explicación del comportamiento¹

Día 1. Teoría de la información

Emmanuel Alcalá

¹Ciclo de charlas auspiciado por la Universidad panamericana, campus Guadalajara y Universidad de Guadalajara-CEIC, con financiamiento CONACyT-320943

Temas

- ① Preliminares
- ② información y entropía

Preliminares

Funciones

Las funciones tienen datos de entrada y datos de salida. Entre la entrada y la salida existe una transformación, que es lo que llamamos “función”. La función $f(x) = y$ se denota por

$$f: X \mapsto Y$$

la función f mapea valores de X (input) a Y (output)

Preliminares

Funciones

Las funciones tienen datos de entrada y datos de salida. Entre la entrada y la salida existe una transformación, que es lo que llamamos “función”. La función $f(x) = y$ se denota por

$$f: X \mapsto Y$$

la función f mapea valores de X (input) a Y (output)

Por ejemplo, la función $y = 5x + 2$ nos dice que por cada valor de x , y valdrá 5 veces x más 2. Si $x = (1, 2, 3)$, la función transforma a x en la colección de valores

$$y = (1 * 5 + 2, 2 * 5 + 2, 3 * 5 + 2)$$

En una tabla es más fácil de visualizar

Operador Σ

Σ - Suma de i elementos hasta n , donde i el índice de sumación

$$\sum_{i=m}^n x_i = x_1 + x_2 + \cdots + x_n$$

Donde m es el límite inferior de la suma, y n el límite superior de la suma.

Operador Σ

Σ - Suma de i elementos hasta n , donde i el índice de sumación

$$\sum_{i=m}^n x_i = x_1 + x_2 + \cdots + x_n$$

Donde m es el límite inferior de la suma, y n el límite superior de la suma.

Ejemplo: sumar los valores de x del 1 al 4

$$x = \{1_{[1]}, 8_{[2]}, 3_{[3]}, 5_{[4]}\}$$

$$\sum_{i=2}^4 = 8 + 3 + 5 = 10$$

En corchete coloco *el orden* (el índice) del elemento

Conjuntos (brevísima)

Un conjunto es una colección de distintos objetos. A es un subconjunto de B si todo elemento de A es también incluido en B , que se simboliza como $A \subset B$.

- El conjunto vacío, denotado \emptyset , es el conjunto que no contiene nada.
- Denotamos por $|S|$ a la cardinalidad (maomeno *el número de sus elementos*) de S . Por ejemplo, $|\emptyset| = 0$.
- $x \in A$ es “ x es un miembro del conjunto A ”. En probabilidad, los eventos se tratan como conjuntos de valores. Si lanzas una moneda 10 veces, el evento “veces que cae águila” es un conjunto de valores.

Probabilidad

Si todos los resultados son igualmente posibles, la probabilidad de A

$$\Pr_{\text{naïve}}(A) = \frac{\text{veces que sale } A}{\text{total de resultados (i.e., } A \cup A^c)} = \frac{|A|}{|S|}$$

Donde por $|\cdot|$ entendemos la *cardinalidad*, o el número de elementos.

Probabilidad

Si todos los resultados son igualmente posibles, la probabilidad de A

$$\Pr_{\text{naïve}}(A) = \frac{\text{veces que sale } A}{\text{total de resultados (i.e., } A \cup A^c)} = \frac{|A|}{|S|}$$

Donde por $|\cdot|$ entendemos la *cardinalidad*, o el número de elementos.

Existen diferentes interpretaciones de la probabilidad. La más usada, la frecuentista, la define formalmente como el valor límite de $\Pr(A)$ cuando $|S|$ tiende a infinito.

Probabilidad

Si todos los resultados son igualmente posibles, la probabilidad de A

$$\Pr_{\text{naïve}}(A) = \frac{\text{veces que sale } A}{\text{total de resultados (i.e., } A \cup A^c)} = \frac{|A|}{|S|}$$

Donde por $|\cdot|$ entendemos la *cardinalidad*, o el número de elementos.

Existen diferentes interpretaciones de la probabilidad. La más usada, la frecuentista, la define formalmente como el valor límite de $\Pr(A)$ cuando $|S|$ tiende a infinito.

Por ejemplo, la probabilidad de obtener caras al lanzar una moneda es el límite de

$$\Pr(\text{caras}) = \frac{\# \text{caras}}{N}$$

cuando el número de lanzamientos $N \rightarrow \infty$.

Variables y espacios:

Variable aleatoria: es una función que mapea los resultados de un experimento aleatorio al conjunto de los números reales (comúnmente). Se suele representar con letra mayúscula (e.g., X).

Existen variables aleatorias discretas (sus valores son finitos o infinitos contables, como $(0, 1, 2, \dots)$), y continuas (sus valores son infinitos no contables).

Variables y espacios:

Variable aleatoria: es una función que mapea los resultados de un experimento aleatorio al conjunto de los números reales (comúnmente). Se suele representar con letra mayúscula (e.g., X).

Existen variables aleatorias discretas (sus valores son finitos o infinitos contables, como $(0, 1, 2, \dots)$), y continuas (sus valores son infinitos no contables).

Espacio muestral: El conjunto de todos los resultados posibles. Se suele representar con Ω . De este conjunto la X mapea a los reales: $X: \Omega \rightarrow \mathbb{R}$. Es decir, a cada elemento de Ω asigna un número real, $X(\omega)$.

Variables y espacios:

Variable aleatoria: es una función que mapea los resultados de un experimento aleatorio al conjunto de los números reales (comúnmente). Se suele representar con letra mayúscula (e.g., X).

Existen variables aleatorias discretas (sus valores son finitos o infinitos contables, como $(0, 1, 2, \dots)$), y continuas (sus valores son infinitos no contables).

Espacio muestral: El conjunto de todos los resultados posibles. Se suele representar con Ω . De este conjunto la X mapea a los reales: $X: \Omega \rightarrow \mathbb{R}$. Es decir, a cada elemento de Ω asigna un número real, $X(\omega)$.

Evento: Subconjunto de Ω , usualmente representado por una vocal mayúscula, e.g., A . Si lanzamos una moneda dos veces, $\Omega = \{HH, HT, TT, TH\}$. El evento "la primera moneda cae H" es $A = \{HH, HT\}$.

Algunas propiedades de \Pr son:

- ① $0 \leq \Pr(A) \leq 1$ y $\sum \Pr(A) = 1$
- ② Si \emptyset es un conjunto nulo, entonces $\Pr(\emptyset) = 0$.
- ③ Si $A \subset B$, entonces $\Pr(A) \leq \Pr(B)$.
- ④ Si A^c denota el complemento de A , entonces $\Pr(A^c) = 1 - \Pr(A)$.
- ⑤ Si $A \cap B = \emptyset$ denota la intersección nula de A y B , entonces $\Pr(A \cup B) = \Pr(A) + \Pr(B)$, \Pr es aditiva para eventos disjuntos.
- ⑥ De otra manera, para eventos arbitrarios A, B , $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- ⑦ Dos eventos A y B son independientes si $\Pr(A \cap B) = \Pr(A) \Pr(B)$.

y_1					
y_2			n_{ij}		
y_3					
	x_1	x_2	x_3	x_4	x_5

Reglas

- 1 $p(x_2, y_1) = \frac{n_{21}}{N}$ (probabilidad conjunta de x_2 y y_1).
- 2 $c_i = \sum_{j=1}^3 n_{ij}$ (suma de todos los n_{ij} con i fijo).
- 3 $p(x_i) = \frac{c_i}{N} = \sum_{j=1}^5 p(x_i, y_j)$ (probabilidad marginal de x_i ; **regla de la suma**).
- 4 $p(y_j|x_i) = \frac{n_{ij}}{c_i}$ (probabilidad condicional de y_j dado que x_i ocurrió).
- 5 $p(x_i, y_i) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \frac{c_i}{N} = p(y_j|x_i)p(x_i)$ (**regla del producto**)

Operador esperanza $E[X]$

Valor esperado - (a.k.a. *media*, *esperanza*, o *promedio*) es una suma ponderada de los posibles resultados de nuestra variable aleatoria. Matemáticamente, si x_1, x_2, x_3, \dots son todos distintos posibles valores que X puede tomar, el valor esperado de X es

$$E[X] = \sum_i x_i p(X = x_i), \text{ si } x \text{ es discreta}$$

$$E[X] = \int_{\mathbb{R}} x f(x) dx, \text{ si } x \text{ es continua}$$

La multiplicación $x_i p(X = x_i)$ es el valor de x_i por la probabilidad de que x_i ocurra. Por brevedad, podemos simplemente escribir $p(x_i)$ para el caso discreto. Para el caso continuo, $f(x)$ denota la función de densidad de probabilidad.

Si tenemos n datos, todos con la misma probabilidad de ser tomados por X , entonces la esperanza es simplemente la media aritmética:

$$E[X] = \frac{1}{n} \sum_i x_i$$

Con $p(x_1) = p(x_2) = \dots = p(x_n) = 1/n$. Por ejemplo, si $X = (1, 5, 9, 10)$, todos con probabilidad $p(x) = 1/4$,

$$E[X] = \frac{1}{n} \sum_i x_i = \frac{1}{4} \times (1 + 5 + 9 + 10)$$

Propiedades de los logaritmos

Los logaritmos solo están definidos para los números reales

$$\log_b(x) = a, \forall x \in \mathbb{R} > 0$$

Que se lee “para todos los x del conjunto \mathbb{R} mayores que 0”.

Un logaritmo se puede definir como el valor al que hay que elevar la base b para obtener x .

Propiedades de los logaritmos

Los logaritmos solo están definidos para los números reales

$$\log_b(x) = a, \forall x \in \mathbb{R} > 0$$

Que se lee “para todos los x del conjunto \mathbb{R} mayores que 0”.

Un logaritmo se puede definir como el valor al que hay que elevar la base b para obtener x .

Por ejemplo, $\log_2(16) = 4$ ('4 es el logaritmo base 2 de 16'). Por lo tanto, para obtener 16 de nuevo elevamos 2 a la cuarta, $2^4 = 16$.

Propiedad 1 $\log_b(x \times y) = \log_b(x) + \log_b(y)$

En palabras, esto significa que el logaritmo base b del producto de dos números es igual a la suma de los logaritmos de esos números.

Propiedad 2 $\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y)$

Esta propiedad es simplemente la operación inversa de la **Propiedad 1**. Esta propiedad permite expresar las razones (o proporciones) en términos de diferencias.

Si, por ejemplo, $x > y$, el rango de valores que puede tomar x/y va desde 1 a infinito. Por otro lado, si $x \leq y$, el rango de valores está entre 0 y 1. Las razones no son funciones simétricas.

Los logaritmos sí. Si $x > y$, $\log_b(x/y) > 1$, si $x < y$, $\log_b(x/y) < 0$. Si $x = y$, $\log_b(x/y) = 0$. El último resultado implica que $\log_b(1) = 0$

Propiedad 3 $\log_b(x^a) = a \log_b(x)$

Esto se sigue de la **Propiedad 1**. Supongamos que $a = 3$, entonces $\log_b(x^a) = \log_b(x^3) = \log_b(x \times x \times x)$, lo que es lo mismo a escribir $\log_b(x) + \log_b(x) + \log_b(x)$, y dado que $\log_b(x)$ se repite 3 veces, la expresión $\log_b(x) + \log_b(x) + \log_b(x)$ es igual a $3 \log_b(x)$.

Propiedad 4 Si $x < y$, $\log_b(x) < \log_b(y)$.

En palabras, esto significa que el $\log_b(x)$ es una función monotónica y estrictamente creciente de x : si x crece, $\log_b(x)$ también crece.

Interludio: ejercicios

- 1 Suponiendo que lanzar dos veces una moneda *justa* son eventos independientes, ¿cuál es la probabilidad de obtener dos caras? Es decir, $p(x_1 = cara \cap x_2 = cara)$. Hint: la probabilidad de una sola cara es $p(cara) = 1/2$.

2

Y	y_1	0.01	0.02	0.03	0.1	0.1
	y_2	0.05	0.1	0.05	0.07	0.2
	y_3	0.1	0.05	0.03	0.05	0.04
		x_1	x_2	x_3	x_4	x_5
		X				

- ¿Cuál es la probabilidad $p(y_1, x_1)$?
- ¿Cuál es la probabilidad $p(y_1)$?
- ¿Cuál es la probabilidad $p(x_1|y_1)$?

Información y entropía

Usualmente, por *información* nos referimos coloquialmente a la cantidad de datos que son guardados, enviados, recibidos o manipulados por algún medio.

Información y entropía

Usualmente, por *información* nos referimos coloquialmente a la cantidad de datos que son guardados, enviados, recibidos o manipulados por algún medio.

Otra forma de concebir la información es como reducción de incertidumbre.

Información y entropía

Usualmente, por *información* nos referimos coloquialmente a la cantidad de datos que son guardados, enviados, recibidos o manipulados por algún medio.

Otra forma de concebir la información es como reducción de incertidumbre.

Por ejemplo, antes de leer un libro su contenido es desconocido. En ese sentido, nuestra incertidumbre es alta. La primera vez que lo leemos, esa incertidumbre con respecto al contenido disminuye y, al mismo tiempo, podríamos decir que ganamos información.

Contenido de información

La noción de información que necesitamos debe tener dos propiedades deseables:

- La información debe ser extensiva y aditiva. La combinación de dos conjuntos de información con la misma cantidad duplican la información.

Contenido de información

La noción de información que necesitamos debe tener dos propiedades deseables:

- La información debe ser extensiva y aditiva. La combinación de dos conjuntos de información con la misma cantidad duplican la información.
- La información reduce la incertidumbre. La cantidad de información es una función monotonica estrictamente decreciente de la incertidumbre de un evento. Cuanto más probable es un evento, menos información conlleva; cuanto menos probable, más información conlleva.

Contenido de información

La noción de información que necesitamos debe tener dos propiedades deseables:

- La información debe ser extensiva y aditiva. La combinación de dos conjuntos de información con la misma cantidad duplican la información.
- La información reduce la incertidumbre. La cantidad de información es una función monotonica estrictamente decreciente de la incertidumbre de un evento. Cuanto más probable es un evento, menos información conlleva; cuanto menos probable, más información conlleva.

Contenido de información

La noción de información que necesitamos debe tener dos propiedades deseables:

- La información debe ser extensiva y aditiva. La combinación de dos conjuntos de información con la misma cantidad duplican la información.
- La información reduce la incertidumbre. La cantidad de información es una función monotonía estrictamente decreciente de la incertidumbre de un evento. Cuanto más probable es un evento, menos información conlleva; cuanto menos probable, más información conlleva.

La única función que satisface esos dos requerimientos es

$$I(x) = -\log_b(p(x)) = \log_b \frac{1}{p(x)}$$

