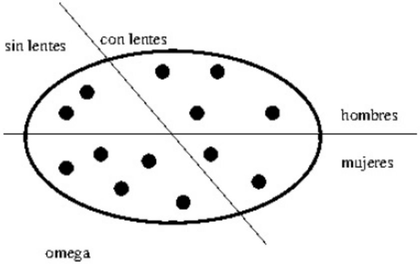
Repaso de probabilidad y estadística II

Alan Reyes-Figueroa

Introducción a la Ciencia de Datos (Aula 03) 17.enero.2022

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

<u>Observaciones</u>:

- $\mathbb{P}(\cdot|B)$ 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 $\mathbb{P}(B|A)$, donde $B = \{primera\ letra\ es\ vocal\}$ y $A = \{\ letra\ es\ b\}$.

Entonces, de la definición de probabilidad condicional, tenemos

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

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

De allí que

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

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

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

<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 55 cartas rojas, y 45 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 $\mathbb{P}(A|E_1) = \dfrac{75}{100}$, $\mathbb{P}(A|E_2) = \dfrac{60}{100}$ y $\mathbb{P}(A|E_3) = \dfrac{45}{100}$.

Entonces, por la ley de probabilidad total

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A|E_1)\,\mathbb{P}(E_1) + \mathbb{P}(A|E_2)\,\mathbb{P}(E_2) + \mathbb{P}(A|E_3)\,\mathbb{P}(E_3) \\
&= \frac{75}{100} \cdot \frac{1}{3} + \frac{60}{100} \cdot \frac{1}{3} + \frac{55}{100} \cdot \frac{1}{3} \\
&= \frac{190}{300} = 0.6333
\end{aligned}$$

UNIVERSIDAD DEL VALLE DE GUATEMALA

Teorema (Regla de Bayes)

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

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

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 que la persona sí tiene Covid, cuando el test da negativo.

Ejemplo

Solución:

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

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

$$
\begin{aligned}
\mathbb{P}(x=1|y=0) &= \frac{\mathbb{P}(y=0|x=1)\,\mathbb{P}(x=1)}{\mathbb{P}(y=0)} \\
&= \frac{(1-\mathbb{P}(y=1|x=1))\,\mathbb{P}(x=1)}{\mathbb{P}(y=0|x=0)\,\mathbb{P}(x=0)+\mathbb{P}(y=0|x=1)\,\mathbb{P}(x=1)} \\
&= \frac{0.03(0.01)}{0.02(0.99)+0.03(0.01)} = 0.01492
\end{aligned}
$$

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

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,*

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

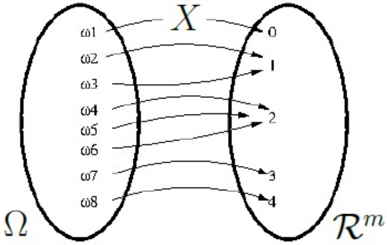
Ejemplo

Lanzamiento de dos dados D_1 y D_2. Consideremos los eventos $A = \{D_1 + D_2 \text{ es par}\}$, $B = \{D_1 < 5\}$, $C = \{D_1 \leq 3, D_2 \leq 3\}$. Sabemos que $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(A \cap B) = \frac{1}{2}$, $\mathbb{P}(A \cap C) = \frac{5}{9}$.

D_1 D_2	1	2	3	4	5	6
1	X		X		X	
2		X		X		X
3	X		X		X	
4		X		X		X
5						
6						

D_1 D_2	1	2	3	4	5	6
1	X		X			
2		X				
3	X		X			
4						
5						
6						

Luego, A y B son independientes; mientras que A y C no lo son.



Definición

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Una ***variable aleatoria*** *(v.a.) es una función mesurable $X : \Omega \to \mathbb{R}$.*

Aquí mesurable significa que si $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, entonces la preimagen de cualquier elemento en $\mathcal{B}(\mathbb{R})$ es un elemento de $\mathcal{F}$. Esto es, X^{-1} lleva conjuntos mesurables de $\mathbb{R}$ (bajo la medida de Lebesgue μ), a conjuntos mesurables en $\mathcal{F}$ (bajo la probabilidad $\mathbb{P}$).

A los elementos de $\mathcal{B}(\mathbb{R})$ se les llama los *borelianos* de $\mathbb{R}$.

Ejemplo
Elegimos al azar una persona de un grupo. De cada persona tenemos un registro de su edad, altura, peso, $\ldots$

Mapeamos cada persona ω a $X(\omega) = (X_1(\omega), \ldots, X_d(\omega))$, donde por ejemplo $X_1(\omega)$ representa su edad, $X_2(\omega)$ su altura, etc.

Si el grupo de personas corresponde a una base de datos, entonces X regresa los campos de interés de cada registro. Las variables $X_1, \ldots, X_d$ son variables aleatorias.
En este ejemplo llamaremos a X como una variable aleatoria (en realidad X es un vector aleatorio).

Observaciones:

- una variable aleatoria determina una relación determinística.
- una variable aleatoria induce una función de probabilidad.

Definimos $\mathbb{P}_X(\cdot)$ como

$$\mathbb{P}_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

Escribimos $\mathbb{P}_X(\cdot)$ como $\mathbb{P}(\cdot)$.
Por ejemplo, $\mathbb{P}(X = x)$ denota $\mathbb{P}_X(X = x) = \mathbb{P}(\{\omega : X(\omega) = x\})$.

$\mathbb{P}(X < a)$ denota $\mathbb{P}_X(X < a) = \mathbb{P}(\omega : X(\omega) < a\})$.

Variables aleatorias

Caso discreto:

Definición

*Diremos que X es una variable aleatoria **discreta** si su contradominio $I = X(\Omega)$ es enumerable y $\mathbb{P}_X(i) = \mathbb{P}(X = i)$ existe para cada $i \in I$. (Comunmente se identifica el contradominio I con los naturales $\mathbb{N}$).*

Definición

*Al conjunto de probabilidades $\{\mathbb{P}_X(i)\}_{i\in I}$ le llamamos la **distribución** de X. (En general, a $\mathbb{P}_X$ se le llama la **función de masa de probabilidad**).*

Definición

*Si $X \in \mathbb{R}$, llamamos a $F_X(x) = \mathbb{P}(X \leq x)$ la **función de distribución** (acumulativa) de X.*

<u>Caso continuo</u>:

Definición

Considere la función $F : \mathbb{R} \to \mathbb{R}$*, dada por*

$$F_X(t) = \mathbb{P}(X \leq t).$$

Diremos que X *es una variable aleatoria* ***continua*** *si existe una función no-negativa* $f_X : \mathbb{R} \to \mathbb{R}$*, tal que*

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(x)\,dx.$$

Definición

En ese caso, a la función f_X *le llamamos la* ***densidad de probabilidad*** *de* X*.*

Propiedades

Obs! La función de densidad f_X no tiene por qué ser continua.

Ya sea en el caso discreto o continuo,

Definición

*Si $x \in \mathbb{R}$, llamamos a $F_X(x) = \mathbb{P}(X \leq x)$ la **función de distribución** (acumulativa) de X.*

En general, definimos la función de distribución para un vector aleatorio $X = (X_1, \ldots, X_d)$ como

$$F_X(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d), \quad \forall (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

En este caso, llamamos a F_X la **función de distribución conjunta** de $X_1, \ldots, X_d$.

Propiedades

Propiedades de $\mathbb{P}_X$ y f_X:

Propiedad	X discreta	X continua
no-negativa	$\mathbb{P}_X(A) \geq 0$	$f_X(x) \geq 0$
suma total	$\sum_x \mathbb{P}_X(x) = 1$	$\int_{\mathbb{R}} f_X(x)\, dx = 1$
relación entre f_X y F_X	$\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$	$f_X(x) = \frac{d}{dx} F_X(x)$
relación entre f_X y F_X	$F_X(x) = \sum_{t \leq x} \mathbb{P}(X = t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$

Propiedades

Propiedades de F_X:

Propiedad	X discreta	X continua
limitada	$0 \leq F_X(x) \leq 1$	$0 \leq F_X(x) \leq 1$
monotonía	F_X no-decreciente	F_X no-decreciente
límite inferior	$F_X(t) = 0, \forall t < \min_{x \in I(\Omega)}$	$\lim_{x \to -\infty} F_X(x) = 0$
límite superior	$F_X(t) = 1, \forall t \geq \max_{x \in I(\Omega)}$	$\lim_{x \to +\infty} F_X(x) = 1$

Además, F_X tiene la propiedad de semi-continuidad inferior: F_X es continua por la derecha, con límites por la izquierda.

Distribuciones conjuntas

Cuando tenemos varias variables aleatorias (definida sobre el mismo espacio $(\Omega, \mathcal{F}, \mathbb{P})$, podemos estudiar la distribución conjunta de dichas variables, esto es, la distribución de (X, Y).

Definición

*La **distribución conjunta** de las v.a. X y Y se define por*

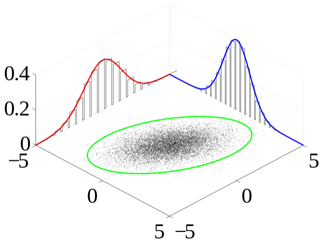
$$F_{X,Y}(X = x, Y = y) = \mathbb{P}(X \leq x, Y \leq y), \;\; \forall\, x, y \in \mathbb{R}.$$

*En el caso que X y Y son v.a. discretas, su **probabilidad conjunta** es*

$$\mathbb{P}_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y), \;\; \forall\, x, y \in \mathbb{R}.$$

*En el caso en que X y Y son continuas, su **densidad conjunta** es*

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial y \partial x}, \;\; \forall\, x, y \in \mathbb{R}.$$



La normal bivariada es la distribución conjunta entre dos normales.

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Distribuciones marginales

Cuando tenemos varias variables aleatorias y su distribución conjunta, podemos “regresar" a las distribuciones originales.

Definición

Dadas X y Y v.a. discretas y su distribución conjunta $\mathbb{P}_{X,Y}$, la ***distribución marginal*** *para X y para Y son*

$$\mathbb{P}_X(x) = \sum_y \mathbb{P}(x,y), \quad \mathbb{P}_Y(y) = \sum_x \mathbb{P}(x,y).$$

En el caso que X y Y son v.a. continuas, y $f_{x,y}$ es su densidad conjunta, la ***densidad marginal*** *de X y de Y son*

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy, \quad f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx.$$

Distribuciones marginales

Ahora, si X, Y toman valores en $[a, b] \times [c, d]$, la **distribución marginal** se calcula como

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_a^x \int_b^y f_{X,Y}(s, t)\, ds\, dt.$$

luego $F_X(x) = F_{X,Y}(x, d), \quad F_Y(y) = F_{X,Y}(b, y)$, y en el caso $b = \infty$ ó $d = \infty$

$$F_X(x) = \lim_{d \to \infty} F_{X,Y}(x, d), \quad F_Y(y) = \lim_{b \to \infty} F_{X,Y}(b, y).$$

Definición

Sean X, Y v.a. discretas tales que $\mathbb{P}(X = x) > 0$. La ***probabilidad condicional*** *de Y dado $X = x$ es*

$$\mathbb{P}_{Y|X}(y|x) = \mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} = \frac{\mathbb{P}_{X,Y}(x, y)}{\mathbb{P}_X(x)}.$$

En el caso continuo, la ***densidad condicional*** *de Y dado X es*

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(y \mid x)}{f_X(x)}.$$

Podemos escribir

- $\mathbb{P}_X(x) = \sum_y \mathbb{P}_{X|Y}(x \mid y)\, \mathbb{P}_Y(y).$
- $f_X(x) = \int_{\mathbb{R}} f_{X|Y}(x \mid y)\, f_Y(y)\, dy.$

Definimos la independencia de variables aleatorias de la siguiente manera:

Definición

Dos variables aleatorias discretas X y Y definidas sobre el mismo espacio Ω son ***independientes*** *si*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y), \;\; \forall\, x, y \in \mathbb{R}.$$

o equivalentemente, $\mathbb{P}_{X,Y} = \mathbb{P}_X \cdot \mathbb{P}_Y$.
En general, las v.a. discretas $X_1, \ldots, X_n$ *son* ***mutuamente independientes*** *si*

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i), \;\; \forall x_1, x_2, \ldots, x_n \in \mathbb{R}.$$

o equivalentemente, $\mathbb{P}_{X_1,\ldots,X_n} = \mathbb{P}_{X_1} \cdot \mathbb{P}_{X_2} \cdot \ldots \cdot \mathbb{P}_{X_n}$.

Definición

Dos variables aleatorias continuas X y Y definidas sobre el mismo espacio Ω son ***independientes*** *si*

$$F_{X,Y}(x,y) = F_X(x)\,F_Y(y), \quad \forall x, y \in \mathbb{R}.$$

En general, las v.a. continuas $X_1, \ldots, X_n$ son ***mutuamente independientes*** *si*

$$F_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \forall x_1, x_2, \ldots, x_n \in \mathbb{R}.$$

Se puede mostrar que esto es equivalente a

$$f_{X,Y}(x,y) = f_X(x)\,f_Y(y), \quad \forall x, y \in \mathbb{R},$$

$$f_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall x_1, x_2, \ldots, x_n \in \mathbb{R}.$$

Referencias

Tarea para el próximo día (pensar, no para entregar).

- Pensar en 2 ejemplos simples de v.a. discretas, y 2 continuas. Para cada uno, graficar su función de masa de probabilida (o densidad), y su función de distribución.
- Pensar en un ejemplo de vector aleatorio discreto y uno continuo. ¿Cómo se ve la función de distribución conjunta en el caso discreto?
- Leer Capítulos 2 y 3 del libro de Lefebvre, hasta la página 60 (antes de comenzar con las distribuciones).
- Se puede leer también los capítulos 1 y 2 en el libro de Chung. El enfoque es mucho más teórico. Les puede servir más para su curso de medida.

Referencia: M. Lefebvre. *Basic Probability Theory with Applications*.