Tema 2. Modelos de distribución multivariados

2.1. Funciones de distribución y densidad

VA, espacios muestrales y realizaciones

X: representa una cantidad desconocida, VA.

$\mathcal{X}$: conjunto de todos los valores posibles que puede tomar la VA. Espacio muestral

Evento: subconjunto de valores del espacio muestral

$\mathcal{X}$ el lado de un dado al lanzarlo

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

Evento: "Cae número par"

$A = \{2, 4, 6\}$

$A \subset \mathcal{X}$

"Cae 2"

$B = \{2\}$

$B \subset \mathcal{X}$

$A = \{X : X \in \{2, 4, 6\}\}$

= Medir la temperatura de un horno

X es una regla, X es una función que retorna un número

$X : \mathcal{X} \longrightarrow \mathbb{R}$

Mapea del espacio muestral al conjunto de los reales

$X(\mathcal{X}) = 75°C$

BMI: índice de masa corporal

$$BMI = \frac{\text{peso en Kg}}{H^2\,(m^2)} \Rightarrow \boxed{X(w, h) = \frac{w}{h^2}}$$

Variable aleatoria como función de dos argumentos

Realización de X

$\underline{\underline{x}}$: realización de X

$\underline{\underline{x}} = 180°C$ $\quad X$

Para BMI

$x = X(70, 1.75) = 22.86 \ kg/m^2$

$p(X = \underline{\underline{x}})$

$p(X = 22.86)$, $p(22.86)$

Definición de probabilidad

Frecuentista

$$\boxed{\lim_{N \to \infty} \frac{|A|}{N} = p(A)}$$

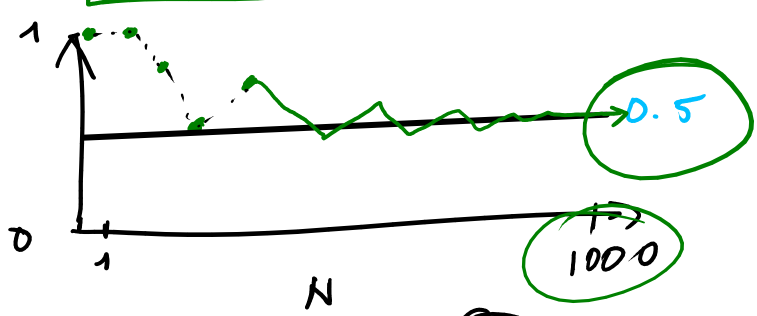
cardinalidad

¿Cuál es la probabilidad de que al lanzar una moneda "caiga" H"?

$\Rightarrow$

$$\boxed{p(X = H) = \frac{1}{2}}$$

$X = \{H, T\}$



$H \Rightarrow \frac{1}{1}, \frac{2}{2}, \frac{2}{3}, \frac{2}{4}, \frac{3}{5}$ ← Frecuencia relativa

$$P(A) = \frac{\text{casos favorables de } A}{\text{total de casos}} = \frac{|A|}{N}$$

$P(\ )$ es una función que mapea a $[0, 1]$

$$P : A \longrightarrow \mathbb{R} \in [0, 1]$$

Algunas propiedades P $\quad A \subset B$ $\Rightarrow$

$A = \{2, 4, 6\}$

$$\boxed{B = \{2, 3, 4, 5, 6\}}$$

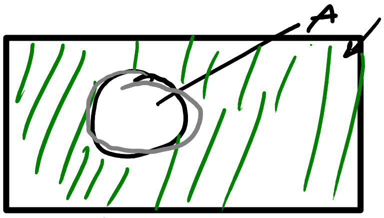
1. $P(\emptyset) = 0$
2. Si $A \subset B$, $P(A) \leq P(B)$
3. A^c indica complemento de A, $P(A^c) = 1 - P(A)$

$P(X=2) = \frac{1}{6}$

$P(X=4) = \frac{1}{6}$ $\Big\}$ $3/6$

$P(X=6) = \frac{1}{6}$

$P(B) = \frac{5}{6}$



$A = \{2, 4, 6\}$

$A^c = \{1, 3, 5\}$

$P(A^c) = 1 - P(A)$

$\Downarrow$

Probabilidad complementaria

"y" "and" $\qquad$ unión

4. Si $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

"intersección de A y B es un evento nulo"

Sea $A = \{1, 3\}$ y sea $B = \{4, 6\}$

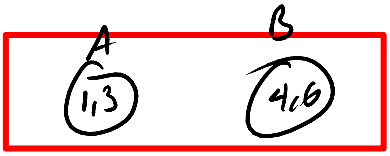
¿Cuál es la probabilidad de que al lanzar moneda, dé un resultado en A o B?

$P(X \in \{1, 3\}) = \frac{1}{6} + \frac{1}{6}$

$P(X \in \{4, 6\}) = \frac{1}{6} + \frac{1}{6}$

$\in$

$P(X \in (1, 3, 4, 6)) = \frac{2}{6} + \frac{2}{6} = \frac{4}{6} = \frac{2}{3}$



$\Leftarrow$ subconjuntos disjuntos (mutuamente excluyentes)

$A = \{1, 3, 4\}$, $B = \{4, 6\}$

$A \cap B = \{4\}$ $\qquad P(X = 4) = \frac{1}{6}$

5. De otra manera $(A \cap B \neq \emptyset)$

$P(A \cup B) = P(A) + P(B) - \boxed{P(A \cap B)}$ $\Rightarrow$ $P(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$

$P(B) = \frac{1}{6} + \frac{1}{6}$

$= \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{2}{3}$

6. Dos eventos A y B son independientes si y solo si

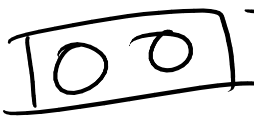
$\Rightarrow$ $P(A \cap B) = P(A)\,P(B)$ $\Leftarrow$

Propiedad empírica

$P(H_1 \cap H_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$

$\Rightarrow$ No confundir independencia con disjunción

$P(A \cap B) = 0$



Reglas de probabilidad

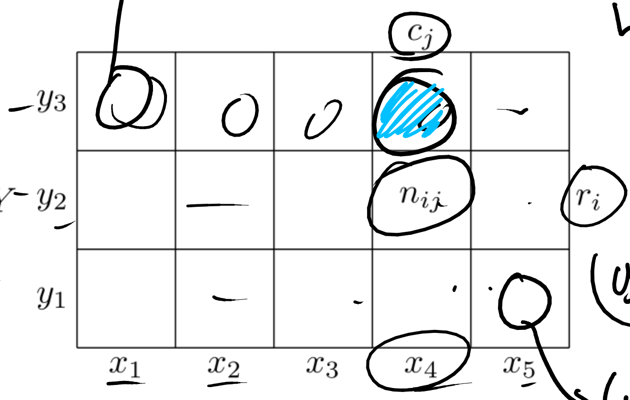
En cada celda coloco la cantidad de personas con una altura dada y un peso dado

Matriz de dos variables bivariadas (conjuntos) Distribución de frecuencias

Altura

$40 \leq y_2 \leq 80$ ↳ Y

$0 \leq y_1 \leq 40$



$(y_3, x_1) = \{80 \leq y_3 \leq 120,\ 0 \leq x_1 \leq 20\}$

$(y_1, x_5) = \{0 \leq y_1 \leq 40,\ 100 \leq x_5 \leq 120\}$

Peso

$0 \leq x_1\ 20$

$20 \leq x_2 \leq 40$

n_{ij}: $i = \{1, 2, 3\}$, $j = \{1, 2, 3, 4, 5\}$

$P(X \cap Y)$

r_2: la cantidad de personas con y_2

c_4: la cantidad de personas con x_4

Probabilidad conjunta

$$P(x, y) \triangleq P(X = x, Y = y)$$

Realizaciones

$$P(X = x_4, Y = y_3) = \frac{n_{34}}{N};$$

$N =$ Sumatoria de la tabla de frecuencias

$$N = \sum_{i=1}^{3} \sum_{j=1}^{5} n_{ij}$$

$$P(X = x_j, Y = y_i) = \frac{n_{ij}}{N}$$

X

[redacted]	0	1	2
0	0	275	416
1	446	2570	1338
2	1251	2523	396
3	462	323	0

Y

$p(y=0)$, $p(y=1)$, $p(y=2)$, $p(y=3)$

Distribución marginal de y

$$P(X=0, Y=2) = \frac{1251}{10000}, \quad N = 10000$$

$$= 0.1251$$

Probabilidad marginal

$$P(Y=2) = p(X=0, Y=2) + p(X=1, Y=2) + p(X=2, Y=2)$$

$$= \frac{1251}{10000} + \frac{2523}{10000} + \frac{396}{10000}$$

$$= \frac{1}{10000}(1251 + 2523 + 396)$$

$$= \frac{4164}{10000} = 0.4164$$

$L = \{1, 2, 3\}$

$$P(Y=y_i) = p(y_i) = \sum_{i=1}^{L} \frac{r_i}{N} = \sum_{j=1}^{M} p(x_j, y_i)$$

i está fija

j variamos en $M = \{0, 1, 2, 3, 4\}$

Regla de la Suma

Fijamos $X = x_1$

Y/X	0	1	2
0	0	275	416
1	446	2570	1338
2	1251	2523	396
3	462	323	0

5691

La fracción de casos en que $Y = 2$

Se escribe como

$$P(Y = y_2 \mid X = x_1) = \frac{n_{21}}{c_1} = \frac{2523}{5691}$$

$= 0.443$

Dado que $X = x_1$, ¿cuál es la probabilidad de que $Y = y_2$?

"N"

275
2570
2523
323

Probabilidad condicional

$$\Rightarrow P(y_i \mid x_j) = \frac{n_{ij}}{c_j}$$

P. marginal $p(x_j)$

$$P(x_j, y_i) = \frac{n_{ij}}{N} = \boxed{\frac{n_{ij}}{c_j}} \cdot \boxed{\frac{c_j}{N}}$$

$p(y_i \mid x_j)$

$p(x_j \mid y_i)$

$$P(x_j, y_i) = p(y_i \mid x_j)\, p(x_j)$$

$$P(y_i, x_j) = \frac{n_{ij}}{r_i} \cdot \frac{r_i}{N} \rightarrow p(y_i)$$

$$p(x_j, y_i) = p(y_i, x_j)$$

$$p(y_i | x_j)\, p(x_j) = p(x_j | y_i)\, p(y_i)$$

$$\boxed{p(y_i | x_j) = \frac{p(x_j | y_i)\, p(y_i)}{p(x_j)}}$$

Teorema de Bayes

$\approx 0.99 \Rightarrow$ VP

FP

X

y

Y/X	0	1	2
0	0	275	416 $\Rightarrow 691$
1	446	2570	1338
2	1251	2523	396
3	462	323	0

1. $P(y=0, x=1) = \frac{275}{10000} \Leftarrow \frac{n_{ij}}{N} = 0.0275$

2. $P(x=1 | y=0) = \frac{n_{ij}}{r_i} = \frac{275}{691} = 0.3979$

3. $F(X \leq 1, y \leq 2)$

Función de distribución acumulada

(cdf: cumulative distribution function)

$$F(X \le 1, Y \le 2) = \sum_{\boxed{x \le a}\ \boxed{y \le b}} p(x, y)$$

$$= p(x=0, y=0) + p(x=0, y=1) + p(x=0, y=2) +$$

$$p(x=1, y=0) + p(x=1, y=1) + p(x=1, y=2)$$

$$= \frac{0}{10000} + \frac{446}{10000} + \frac{1251}{10000} + \frac{275}{10000} + \frac{2570}{10000}$$

$$+ \frac{2523}{10000} = \frac{7065}{10000} = 0.7065$$

$F(Y \le 3)$

Y/X	0	1	2
0	0	275	416
1	446	2570	1338
2	1251	2523	396
3	462	323	0

$\frac{0}{10000}$

446/10000

1251/10000

462/10000

N/N

Distribución $P(X_0)$

marginal

$\Sigma 1$

$$\stackrel{?}{=} \sum_{j=0}^{2} \boxed{\sum_{i}^{M} p(x_j, y_i)}$$

$x \le a$

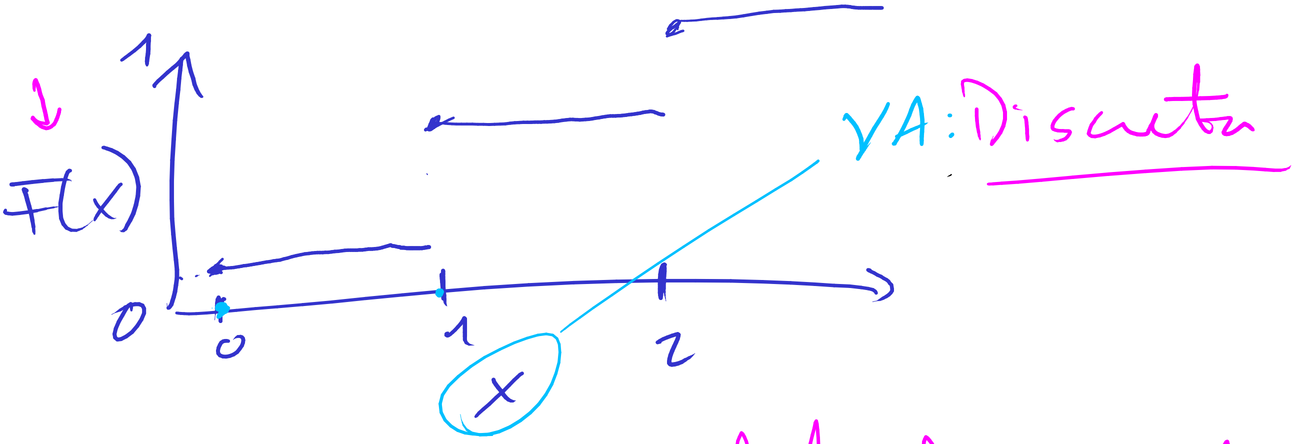
$$P(X_j) = \sum_{i}^{M} p(x_j, y_i)$$

$$F(X \le 2) = p(X_0) + p(X_1) + p(X_2)$$

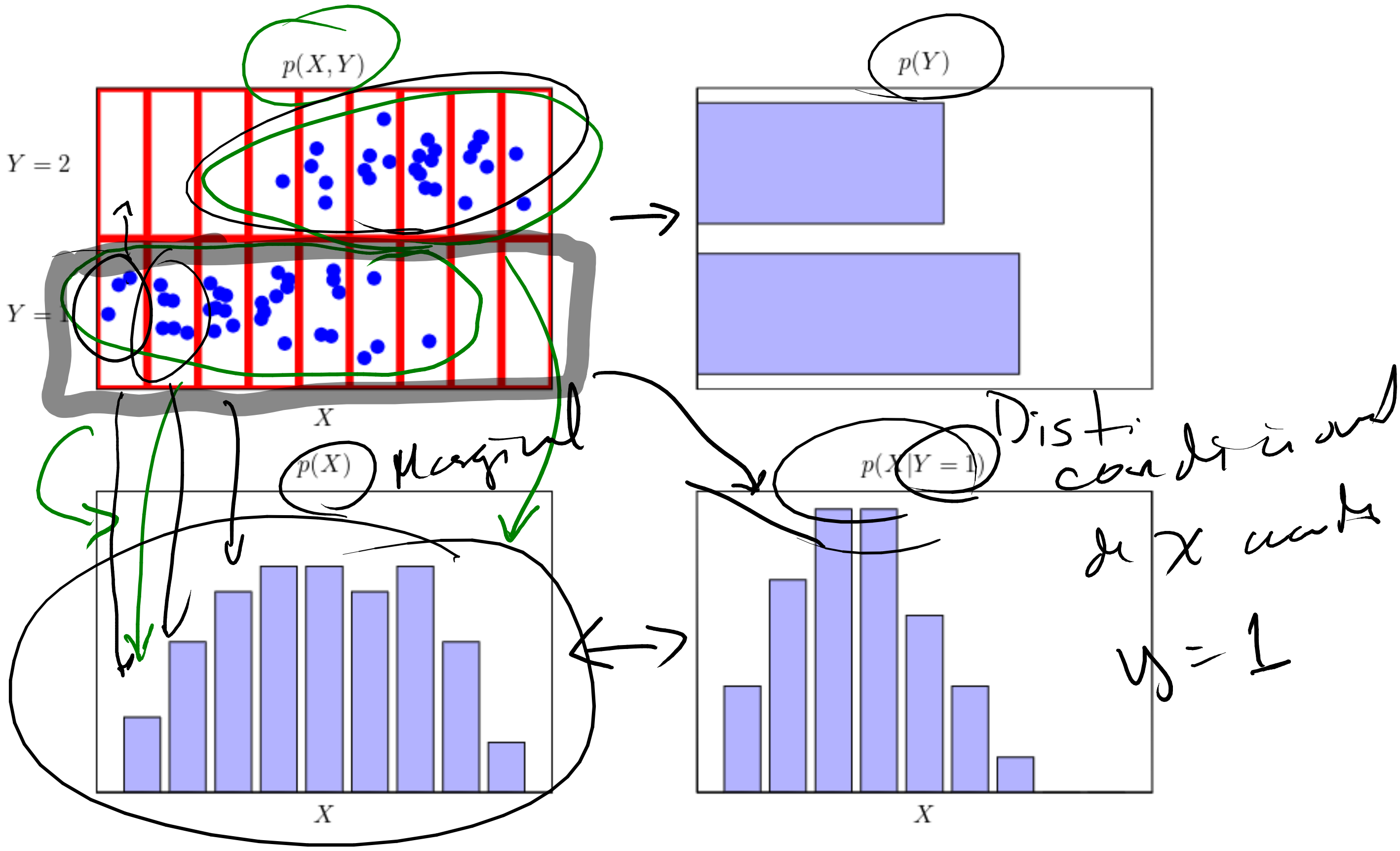
$$= \sum_{i=0}^{M} p(X_0, y_i) + \sum_{i=1}^{M} p(X_1, y_i) +$$

$$+ \sum_{i=1}^{M} p(X_2, y_i)$$

$$F(X \le 2) = 0.2159 + 0.5691 + 0.2150 = 1$$



La dist. acumulada de una VA. discreta es discontinua.



17 – 50

$p(X \mid 17 \leq y \leq 50)$

↳ Restricción de rango

$P(x)$

Para VA discretas

$P(x)$ lo obtenemos una función de probabilidad; función de masa de probabilidad (pmf: probability mass function)

Si X es VA discreta

Para VA continua, $X \in \mathbb{R}$

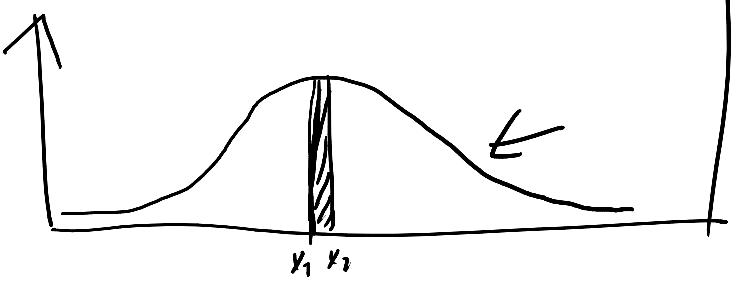
$f(x)$ La función de densidad de probabilidad (pdf: probability density function)

$P(X=2) \leftarrow$ si está definido si X es VA discreta.

$f(X=2)=0$

$\Delta x \doteq X_2 - X_1$

Para VA cont. especificamos un intervalo en donde se encuentre x.



$$F(a \le x \le b) = \int_a^b f(x)\,dx$$

$f(x)$ es continua

$$F(0 \le x \le 0) = \int_0^0 f(x)\,dx = P(x)\Big|_0^0 = P(x=0) - P(x=0) = 0$$

Función marginal de x

Discreta

Regla de la suma

$$p(x) = \sum_y p(x, y) \;; \quad p(x) = \int_y f(x,y)\,dy$$

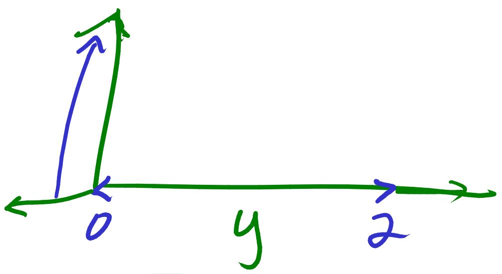
$$p(x, y) = p(y|x)\, p(x)$$

Ejemplos con $\underbrace{\text{funciones conocidas}}$

forma funcional (tienen una ecuación)

1. Dada $f(y) = cy^2$, $0 \le y \le 2$, y $f(y) = 0$ en cualquier otra parte, encuentra el valor de C para el cual $f(y)$ es una función de densidad válida.

$$f(y) = \begin{cases} cy^2 & \text{si } 0 \le y \le 2 \\ 0 & \text{en cualquier otro caso} \end{cases}$$



Dos propiedades de cualquier función de densidad

1. $f(y) \ge 0$ → 1. $p(y) \ge 0$
2. $\boxed{\int_{-\infty}^{\infty} f(y)\,dy = 1}$ → 2. $\sum_y p(y) = 1$

$$F(y) = \int_{-\infty}^{\infty} f(y)\,dy = 1$$

cdf

1. Obtener la integral
2. Despejas para c

$$\int_0^b x^n\,dx = \left.\frac{x^{n+1}}{\boxed{n+1}}\right|_0^b$$

$$\Downarrow$$

$$\int_0^b k x^n\,dx = k\int_0^b x^n\,dx$$

$$F(y) = \int_0^2 c y^2\,dy = 1$$

$$= c\int_0^2 y^2\,dy = c\left[\frac{y^3}{3}\right]_0^2 = c\left[\frac{(2)^3}{3} - \frac{(0)^3}{3}\right]$$

$$= c\left[\frac{8}{3}\right]\left(\frac{3}{8}\right) = 1\left(\frac{3}{8}\right) \Rightarrow \boxed{c = \frac{3}{8}}$$

2. Encontrar $F(1 \leq y \leq 2)$ con $f(y) = \frac{3}{8}y^2$
(también ~~encontrar~~ $F(1 < y < 2)$).

$$F(1 \leq y \leq 2) = \int_1^2 \frac{3}{8}y^2\,dy = \frac{3}{8}\int_1^2 y^2\,dy = \frac{3}{8}\left[\frac{y^3}{3}\right]_1^2$$

$$= \frac{3}{8}\left[\frac{(2)^3}{3} - \frac{(1)^3}{3}\right] = \frac{\cancel{3}}{8}\left(\frac{7}{\cancel{3}}\right)$$

$$= \frac{7}{8}$$

$$\boxed{F(a \leq y \leq b) = F(a < y < b)} + F(a = y = b)$$

0

3. Como medida de inteligencia, a unos ratones se les toma tiempo que tardan en pasar por un laberinto.
El tiempo en segundos necesario para cualquier ratón es una VA Y con pdf dada por

$$f(y) = \begin{cases} \frac{b}{y^2}, & \text{si } y \geq b \\ 0 & \text{en cualquier otro punto} \end{cases}$$

En donde b es el tiempo mínimo posible para recorrer el laberinto.
Demostrar que f(y) es una pdf válida y encontrar para qué valores de b es válida

b al ser tiempo es $b \geq 0$, y^2

$f(y) \geq 0$ ✓

$\Rightarrow$ 1. $f(y) \geq 0$

$\nRightarrow$ 2. $F(y) = \int_{-\infty}^{\infty} f(y)\,dy = 1$

$\int_b^{?}$