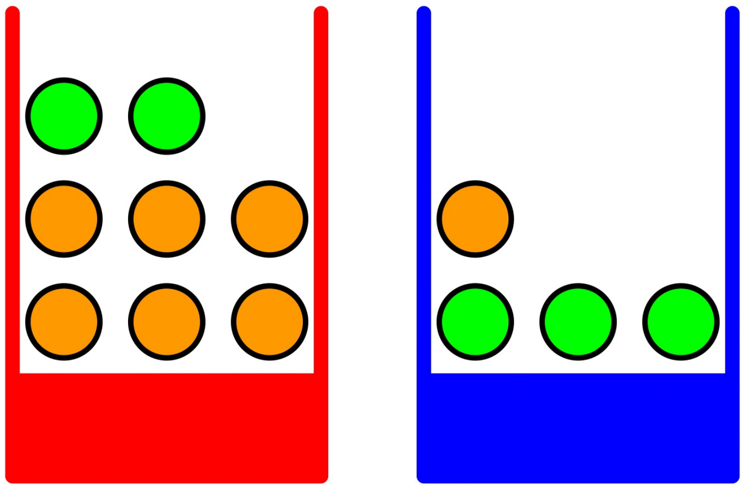
Repaso Probabilidad

$B \in \{r, b\}$; CAJAS

MANZANA NARANJA.

X: variable aleatoria CAJA

Y: variable aleatoria FRUTA

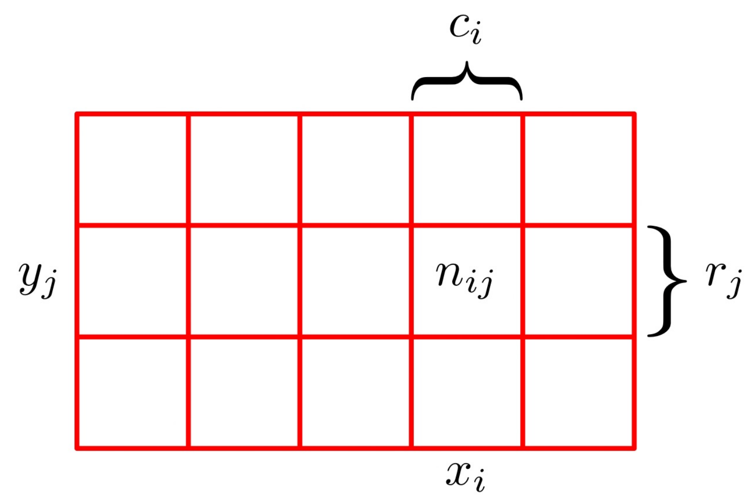


CAJA r → 40% CHANCES CAJA ROJA

$p(B=r) = 0.4$

CAJA b → 60% CHANCES CAJA AZUL

$p(B=b) = 0.6$

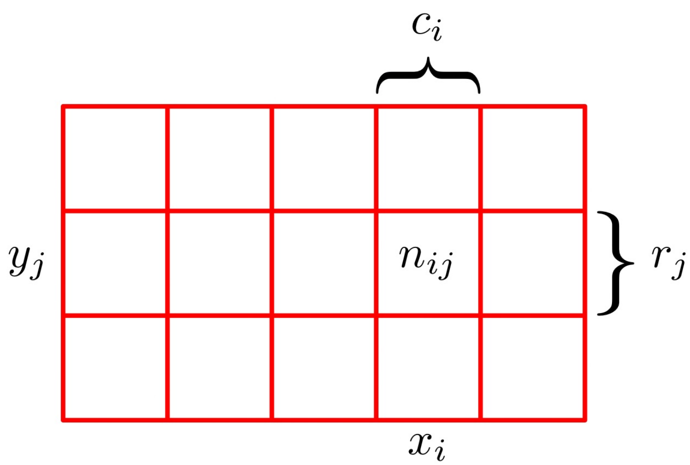


$$p(X=x_i, Y=y_j) = \frac{n_{ij}}{N}$$

$$p(X=x_i) = \frac{c_i}{N}; \quad c_i = \sum_j n_{ij}$$

$$p(Y=y_j) = \frac{r_j}{N}; \quad r_j = \sum_i n_{ij}$$

N MUESTRAS



$$\underbrace{\underset{\text{MARGINAL}}{\underset{\downarrow}{p(X=x_i)}} = \sum_{j=1}^{L} \underset{\text{CONJUNTA}}{\underbrace{p(X=x_i, Y=y_j)}}}_{\text{REGLA DE LA SUMA DE PROBABILIDADES}}$$

$$\underset{\text{CONDICIONAL}}{\underset{\downarrow}{p(Y=y_j \mid X=x_i)}} = \frac{n_{ij}}{c_i}$$

$$\underbrace{p(X=x_i, Y=y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i}\frac{c_i}{N} = p(Y=y_j \mid X=x_i)\, p(X=x_i)}_{\text{REGLA DEL PRODUCTO.}}$$

REGLAS DE PROBABILIDAD:

SUMA: $p(X) = \sum_Y p(X,Y)$ → MARGINALIZAR

PRODUCTO: $p(X,Y) = p(Y|X)p(X) = p(X|Y)p(Y)$

$p(X,Y) = p(Y,X)$ → SIMETRIA

TEOREMA DE BAYES:

$$p(Y|X)p(X) = p(X|Y)p(Y)$$

VEROSIMILITUD

$$p(Y|X) = \frac{p(X|Y)\,p(Y)}{p(X)}$$

PRIOR

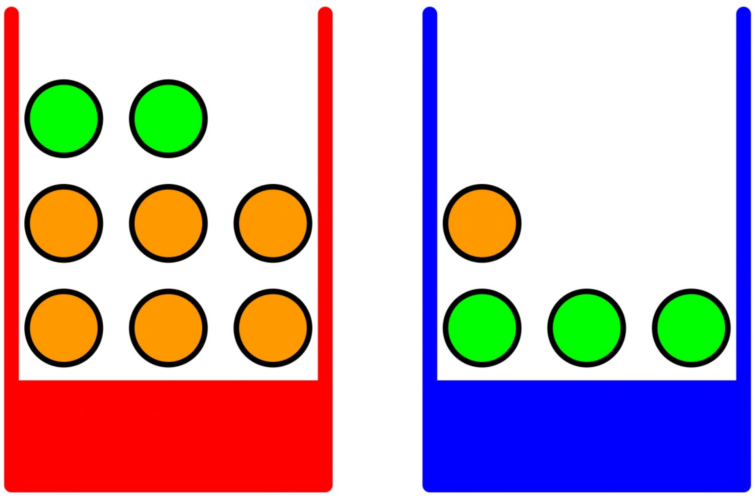
POSTERIOR

EVIDENCIA.

$$p(X) = \sum_Y p(X,Y) = \sum_Y p(X|Y)p(Y)$$

EJERCICIO: PARA $p(B=r)=0.4$ y $p(B=b)=0.6$, $p(B=r)+p(B=b)=1$. y

según la figura, encontrar:



$p(F=a)=?$

$p(F=o)=?$

$p(B=r|F=o)=?$

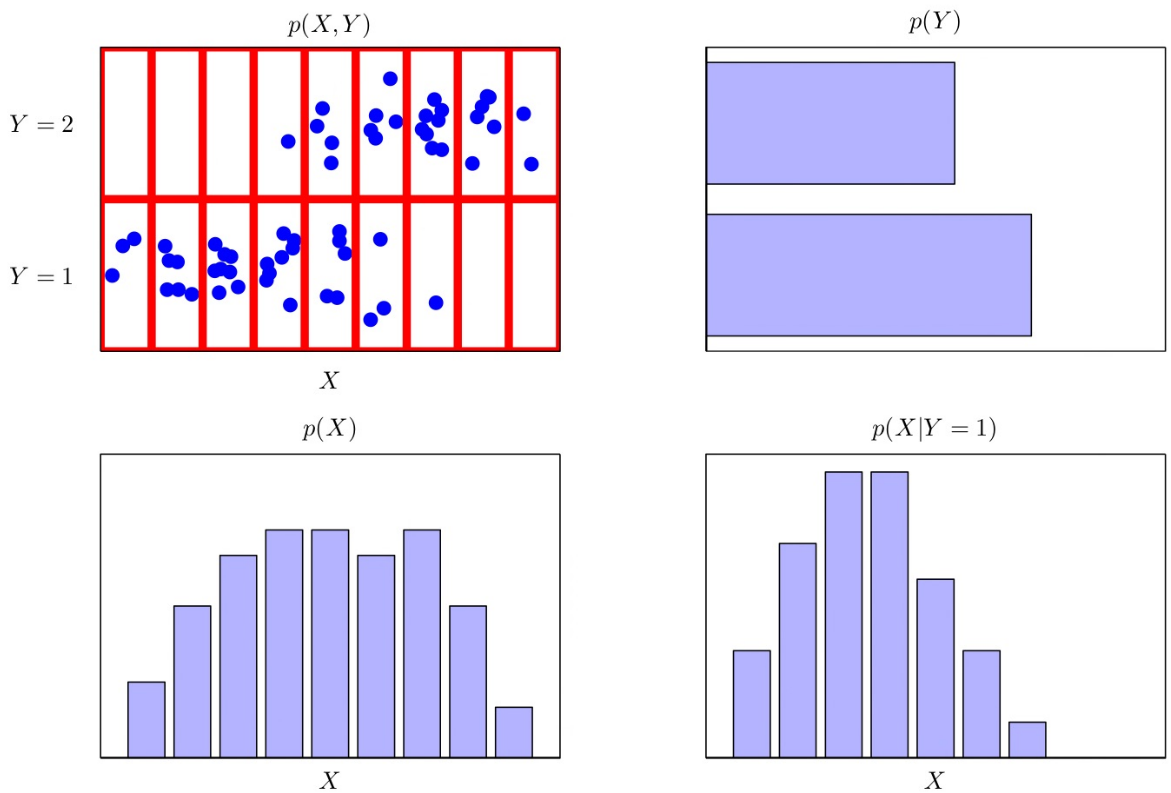
$p(B=b|F=o)=?$

$p(B=r|F=a)=?$

$p(B=b|F=a)=?$

a: apple
o: orange.
B: Box
F: Fruit.

EJERCICIO: Genere simulación en Python que emule las siguientes Figs:

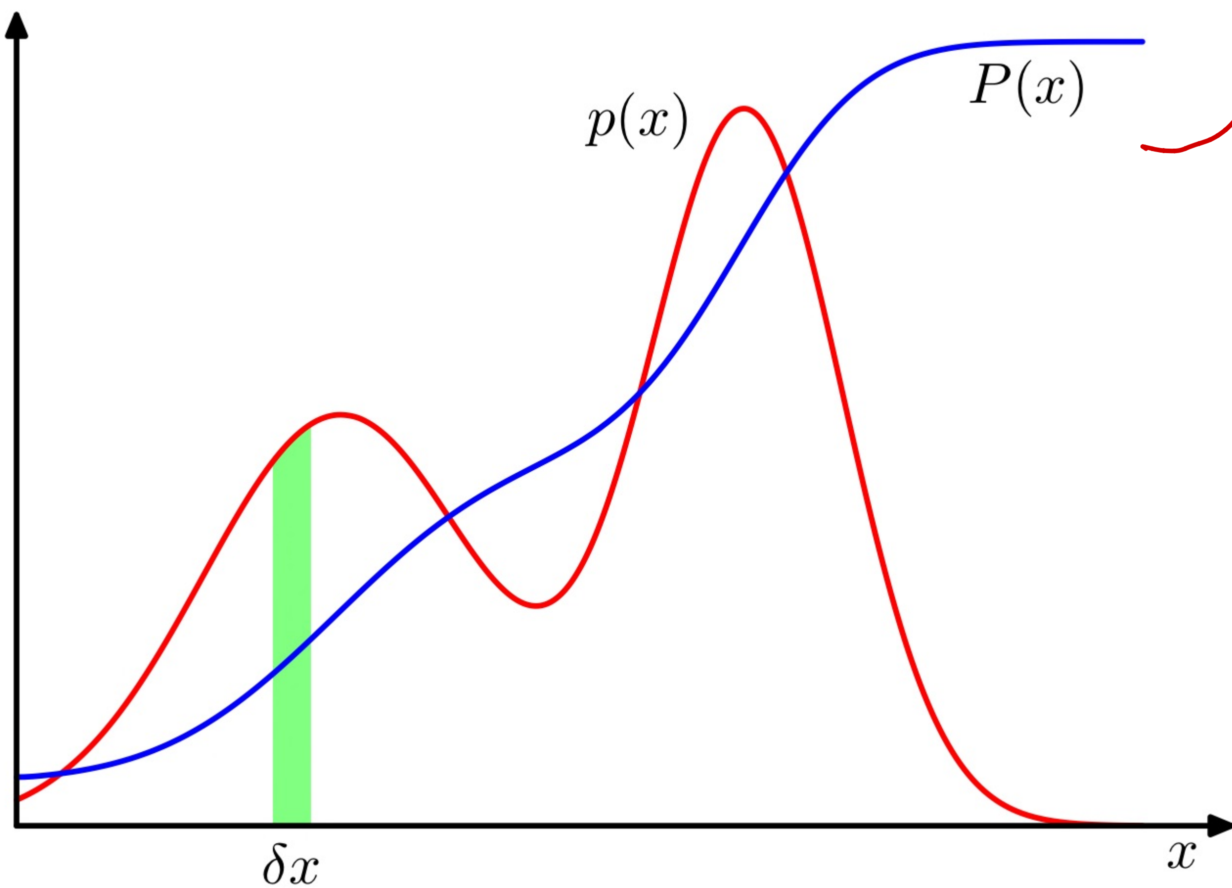


NOTA: Asuma $N=100$ y $p(x|Y=y) = \mathcal{N}(x|\mu_y, \sigma_y^2)$

Funciones Densidad de Probabilidad.

$x \in \mathbb{R}$; la probabilidad de $x \in [x, x+\delta x]$ es $p(x)\delta x$ con $\delta x \to 0$ $\therefore$ $p(x)$ es la densidad de probabilidad sobre x:

$$p(x \in [a,b]) = \int_a^b p(x)\,dx$$



Función de distribución Acumulada

$$P(z) = \int_{-\infty}^{z} p(x)\,dx$$

$$P'(x) = p(x)$$

Consideraciones:

$$p(x) \geq 0 \qquad \int_{-\infty}^{\infty} p(x)\,\partial x = 1$$

Cambio de variable:

$$x = g(y) \quad \therefore \quad f(x) \Rightarrow \tilde{f}(y) = f(g(y))$$

Asumiendo que: $p_x(x)\,\delta x \simeq p_y(y)\,\partial y$

Entonces:

$$p_y(y) = p_x(x)\left|\frac{\partial x}{\partial y}\right|$$

$$p_y(y) = p_x(x)\,|g'(y)|$$

Se asume que para $x \in [x, x+\delta x]$ existe una transformación $y \in [y, x+\delta y]$

Extensión vectorial:

Para $x_1, x_2, \dots, x_D$ variables aleatorias:

$$\mathbf{x} = [x_1, x_2, \dots, x_D] \Rightarrow p(\mathbf{x}) = p(x_1, x_2, \dots, x_D)$$

$$p(\mathbf{x}) \geq 0$$

$$\int_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \, \partial \mathbf{x} = 1.$$

Regla de la suma y producto para $x, y \in \mathbb{R}$

$$p(x) = \int_y p(x, y) \, \partial y$$

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

Operador Esperanza y Covarianzas

$$\mathcal{E}\{f\} = \underbrace{\sum_x p(x) f(x)}_{\text{suma ponderada}}$$

Caso continuo:

$$\mathcal{E}\{f\} = \int_x p(x) f(x)\, dx$$

Media Muestral: para N pontos:

$$\mathcal{E}\{f\} \simeq \underbrace{\frac{1}{N} \sum_n f(x_n)}_{\text{Estimador (Aproxima } \mathcal{E}\{f\})}$$

↳ Aproximación converge para $N \to \infty$

Operador Esperanza en varias variables

$$\mathcal{E}_x\{f(x,y)\} = \int p(x)f(x,y)\,\partial x$$

$$\mathcal{E}_{x,y}\{f(x,y)\} = \int p(x,y)f(x,y)\,\partial xy$$

Esperanza Condicional:

$$\mathcal{E}_x\{f|y\} = \int_x p(x|y)f(x)\,\partial x$$

Momentos básicos:

$$\mu_f = \mathcal{E}_x\{f\} = \int p(x)f(x)\,\partial x$$

$$\mathrm{Var}\{f\} = \mathcal{E}\{(f(x) - \mathcal{E}_x\{f\})^2\}$$

EJERCICIO:

Demuestre que:

$$\text{var}\{x\} = \mathcal{E}\{(x-\mu_x)^2\} = \mathcal{E}\{x^2\} - \mathcal{E}^2\{x\}$$

$$\text{cov}\{x,y\} = \mathcal{E}_{x,y}\{(x-\mu_x)(y-\mu_y)\}$$

$$= \mathcal{E}_{x,y}\{xy\} - \mathcal{E}\{x\}\mathcal{E}\{y\}$$

$$\text{cov}(\mathbb{x},\mathbb{y}) = \mathcal{E}_{\mathbb{x},\mathbb{y}}\{\mathbb{x}\mathbb{y}^T\} - \mathcal{E}\{\mathbb{x}\}\mathcal{E}\{\mathbb{y}\}$$

$$\mathbb{x},\mathbb{y} \in \mathbb{R}^{D\times 1}$$