

Soluciones examen de PyE julio 2018

1. X = número de caras de Ana, Y = número de caras de Beto, $Z = \max\{X, Y\}$.

a) La distribución conjunta de X e Y es

$X \setminus Y$	0	1	2	Marg. X
0	1/16	1/8	1/16	1/4
1	1/8	1/4	1/8	1/2
2	1/16	1/8	1/16	1/4
Marg. Y	1/4	1/2	1/4	1

De aquí se deduce que $\mathbf{P}(X = Y) = \frac{1}{16} + \frac{1}{4} + \frac{1}{16} = \frac{3}{8}$.

b) La distribución conjunta de X y Z es

$X \setminus Z$	0	1	2	Marg. X
0	1/16	1/8	1/16	1/4
1	0	3/8	1/8	1/2
2	0	0	1/4	1/4
Marg. Z	1/16	1/2	7/16	1

No son independientes pues, por ejemplo $\mathbf{P}(X = 2, Z = 0) = 0 \neq \mathbf{P}(X = 2) \mathbf{P}(Z = 0) = 1/4 \times 1/16$.

c) La covarianza es $\mathbf{cov}(X, Z) = \mathbf{E}(XZ) - \mathbf{E}(X) \mathbf{E}(Z)$, y

$$\mathbf{E}(XZ) = 1 \times 1 \times 3/8 + 1 \times 2 \times 1/8 + 2 \times 2 \times 1/4 = 13/8,$$

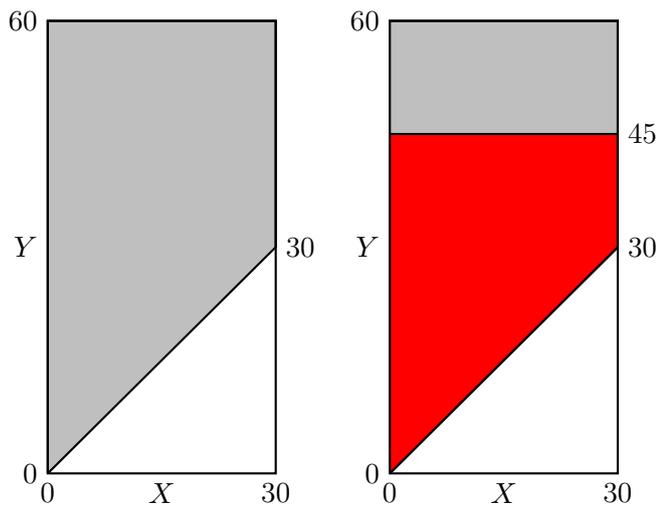
$$\mathbf{E}(X) = 1 \times 1/2 + 2 \times 1/4 = 1,$$

$$\mathbf{E}(Z) = 1 \times 1/2 + 2 \times 7/16 = 11/8.$$

Entonces $\mathbf{cov}(X, Z) = 13/8 - 1 \times 11/8 = 1/4$.

2. Sean X = la hora de llegada de Beto e Y = la hora de llegada de Ana.

a) Del dibujo se ve que el evento “Beto llega antes que Ana” tiene probabilidad $3/4$.



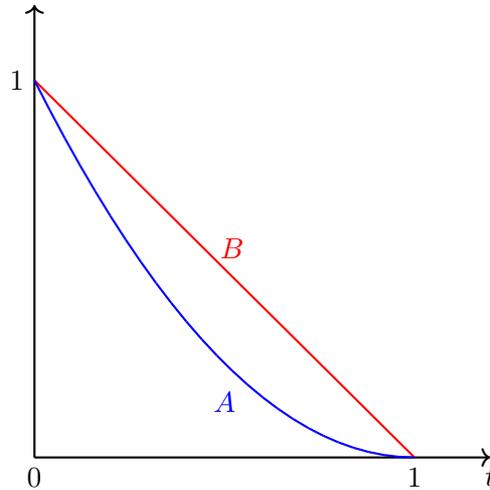
a)

b)

b) La probabilidad del evento rojo (se encuentran) dado que Beto llega antes es $(3/4 - 1/4)/(3/4) = 1 - 1/3 = 2/3$.

3. Para que el área debajo de la densidad sea igual a 1 la altura h del triángulo debe satisfacer $h \times 1/2 = 1$, es decir que $h = 2$. Luego, la densidad de T_A es $f_A(x) = 2(1 - x)$.

a) La probabilidad $\mathbf{P}(T_A \geq t) = (1 - t)^2$, pues es el área de un triángulo de base $1 - t$ y altura $2(1 - t)$. La probabilidad $\mathbf{P}(T_B \geq t) = 1 - t$. Las gráficas son:



Como $\mathbf{P}(T_A \geq t) \leq \mathbf{P}(T_B \geq t)$ para todo $t \in [0, 1]$, se debe elegir el mostrador A .

b) La esperanza de $\mathbf{E}(T_B) = 1/2$. Por otro lado $\mathbf{E}(T_A) = \int_0^1 x \cdot 2(1 - x)dx = 1/3$.

4. El valor observado de X es $X_{\text{obs}} = 484 - 477 = 7$.

a) De la figura deducimos que el p-valor es

$$\text{pval}(7) = \mathbf{P}(X \geq 7) = \frac{3257}{12870} = 0,253$$

y como es mayor que 0,05 no se rechaza H_0 .

b) Como $\alpha \times 12870 = 643,5$, debemos buscar el menor valor de c tal que la suma de las barras es menor o igual a 643. Para $c = 15$ la suma es 570, y para $c = 13$ la suma es 1085. Luego el menor valor es $c = 15$. Como $X_{\text{obs}} = 7 \notin [15, +\infty)$, no se rechaza H_0 .

5. Dado que $X = 60$, la Y tiene distribución normal de parámetros

$$\mu = \rho \frac{\sigma_Y}{\sigma_X} (60 - \mu_X) + \mu_Y = 145 \text{ y } \sigma = \sigma_Y \sqrt{1 - \rho^2} = 20 \frac{\sqrt{7}}{4} = 13,23.$$

a) Entonces

$$\mathbf{P}(Y \geq y | X = 60) = \mathbf{P}\left(\frac{Y - 145}{13,23} \geq \frac{y - 145}{13,23}\right) = 0,05,$$

por lo que $\frac{y-145}{13,23} = 1,645$, de donde $y = 166,76$.

b) Basta encontrar y' tal que $\mathbf{P}(Y \leq y' | X = 60) = 0,05$. Los mismos cálculos que en la parte a) dan $y = 145 - 1,645 \cdot 13,23 = 123,24$. Por lo tanto un intervalo que sirve es $[123,24, 166,76]$.

6. a) La función de verosimilitud es

$$L(a) = \prod_{i=1}^n (a + 1) X_i^a$$

Tomando logaritmos

$$\ell(a) = n \log(a + 1) + a \sum_{i=1}^n \log X_i$$

y la derivada es

$$\ell'(a) = \frac{n}{a + 1} + \sum_{i=1}^n \log X_i$$

Igualando a cero y despejando:

$$\hat{a} + 1 = \frac{-1}{\frac{\sum_{i=1}^n \log X_i}{n}} \Rightarrow \hat{a} = -1 - \frac{1}{\frac{\sum_{i=1}^n \log X_i}{n}}$$

- b) Por la LGN tenemos que $\frac{1}{n} \sum_{i=1}^n \log X_i$ es un estimador consistente de $\mathbf{E}(\log X)$. Por la propiedad de continuidad de los estimadores consistentes, \hat{a} es un estimador consistente de $\frac{-1}{\mathbf{E}(\log X)} - 1$. Utilizando el dato dado en la letra se ve que $\frac{-1}{\mathbf{E}(\log X)} - 1 = a$, por lo que \hat{a} es un estimador consistente de a .

7. Aplicaremos la prueba de comparación de Kolmogorov-Smirnov para ver si ambas muestras tienen la misma distribución. En este caso $n = 6$ y $m = 8$.

- Si d_1 es el valor más grande de $|\frac{i}{n} - F_m^Y(x_i^*)|$, entonces $d_1 = 0,333$
- Si d_2 es el valor más grande de $|\frac{j}{m} - F_n^X(y_j^*)|$, entonces $d_2 = 0,208$

Luego, $d_{n,m} = 0,333$. Según la tabla de KS para dos muestras ($n = 6$ y $m = 8$):

$$\begin{aligned} p - \text{valor} &= \mathbb{P}(D_{n,m} \geq 0,333 | H_0) = \mathbb{P}(nmD_{n,m} \geq nm0,333 | H_0) = \mathbb{P}(nmD_{n,m} \geq 16 | H_0) \\ &> \mathbb{P}(nmD_{n,m} \geq 28 | H_0) = 0,139 \end{aligned}$$

Entonces el p-valor es mayor que 0,139, por lo tanto NO rechazamos H_0 a nivel $\alpha = 0,05$.

8. En un estudio reciente que involucró una muestra de 300 accidentes automovilísticos, se clasificó la información de acuerdo al tamaño del auto y si hubo o no algún muerto como consecuencia del accidente. Los resultados fueron los siguientes:

Frecuencias observadas				
	Pequeño (P)	Mediano (M)	Grande (G)	Total
Al menos un muerto (m)	42	35	20	97
Ningún muerto (n)	78	65	60	203
Total	120	100	80	300

Se quiere testear la hipótesis de que la ocurrencia de muertes en un accidente de tránsito es independiente del tamaño del vehículo.

a) **Frecuencias esperadas si H_0 es cierta.**

	Pequeño (P)	Mediano (M)	Grande (G)	Total
Al menos un muerto (m)	38	33	26	97
Ningún muerto (n)	82	67	54	203
Total	120	100	80	300

b)

$$\chi_{obs}^2 := \sum_{i \in \{m,n\}, j \in \{P,M,G\}} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 2,846$$

- c) El p-valor es $p - \text{valor} = 0,2$, por lo que no se puede rechazar H_0 .