Ciencia de Datos 2021

Lista 01b

30.enero.2021

1. Sean X, Y variables aleatorias. Demostrar las siguientes propiedades:
   a) Si $\mathbb{V}ar(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.
   b) Si X y Y son independientes, entonces $Cov(X, Y) = 0$.
   c) Mostrar que $I(X, Y) = I(Y, X)$.
   d) Proporciona un ejemplo donde no se cumple $\mathbb{V}ar(X + Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y)$.

2. Sean X, Y variables con las siguiente función de probabilidad conjunta:

$\mathbb{P}(x, y)$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$y = 1$	$1/8$	$1/16$	$1/32$	$1/32$
$y = 2$	$1/16$	$1/8$	$1/32$	$1/32$
$y = 3$	$1/16$	$1/8$	$1/32$	$1/32$
$y = 4$	$1/16$	$1/16$	$1/16$	$1/16$

   Calcular
   a) $H(X)$ y $H(Y)$,
   b) $H_Y(X)$ y $H_X(Y)$,
   c) $H(X, Y)$,
   d) $I(X, Y)$.

3. Sean X, Y variables aleatorias conjuntamente continuas con función de densidad

$$f(x, y) = \begin{cases} 2, & \text{si } y + x \leq 1, x > 0, y > 0; \\ 0, & \text{en otro caso.} \end{cases}$$

   Encontrar $Cov(X, Y)$ y $\rho(X, Y)$ (correlación).

4. En el lenguaje de programación de su preferencia (Python, R, Matlab, C^{++}, o cualquier otro que quieran), hacer lo siguiente:
   a) Construir una muestra de 200 elementos, de una variable aletoria $X \in \mathbb{R}^3$ que sigue una distribución normal $\mathcal{N}_3(\mu, \Sigma)$, donde $\mu \in \mathbb{R}^3$ y $\Sigma \in \mathbb{R}^{3\times3}$ son de su elección (Σ debe ser positiva definida, y tener valores 1 en su diagonal principal).
   b) Hacer un plot en 3D de la nube de puntos de esta distribución.
   c) Hacer un pairplot de esta distribución.

   Repita lo anterior para diferentes matrices Σ (3 veces). En función de las gráficas obtenidas, comente sus hallazgos.

5. Descargue la base de datos Iris (publicada orignalmente por Ronald Fisher en 1936).
   `https://archive.ics.uci.edu/ml/datasets/iris`
   `https://en.wikipedia.org/wiki/Iris_flower_data_set`

   Hacer un análisis exploratorio de estos datos (*e.g.* plot de los datos, histogramas, graficar distribuciones univariadas, pairplots, calcular resúmenes de distribuciones). A partir de esta información, ¿qué puede concluir de sus datos?

6. (Ley de probabilidad total para esperanzas, caso finito).
   Sean X y Y variables aleatorias finitas. Mostrar que

$$\mathbb{E}(X) = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y = y]\, \mathbb{P}(Y = y).$$

   ¿Qué ocurre si X o Y son discretas, pero alguna de ellas ya no es finita?

7. Para las siguientes variables aleatorias, calcular $\mathbb{E}(X)$, $\mathbb{V}ar(X)$, $I(X)$, f_X e F_X. (Sugerencia: Hacer una tabla, con todas las distribuciones que vayan viendo en el curso).
   a) $X \sim Unif[1, n]$, donde $[1, n] = \{1, 2, 3, \ldots, n\}$.
   b) $X \sim Ber(p)$.
   c) $X \sim Binom(n, p)$.
   d) $X \sim Geom(\theta)$, $0 < \theta < 1$.
   e) $X \sim Possion(\lambda)$, $\lambda > 0$.

8. Para las siguientes variables aleatorias continuas, calcular $\mathbb{E}(X)$, $\mathbb{V}ar(X)$, $I(X)$, f_X e F_X. (Sugerencia: Hacer una tabla).
   a) $X \sim U(a, b)$, $a < b$.
   b) $X \sim Exp(\lambda)$, $\lambda > 0$.
   c) $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$.