Tarea 1

Vectores aleatorios y transformaciones
Fecha de entrega: martes 31 de octubre del 2023

Instrucciones: Responde las siguientes preguntas y justifica con los procedimientos correspondientes.

Importante: La tarea deberá cumplir con los siguientes requisitos para tener derecho a ser calificado: 1) nombres de los integrantes del equipo (máximo 3 alumnos) en la parte superior derecha de la primera hoja sin ocupar mucho espacio, 2) tener excelente presentación, 3) sin desperdiciar espacios, 4) gráficas bien diseñadas, cuidando escalas y espacios y 5) cuando corresponda justificando con prosa de forma concisa, bien pensada y escrita sus respuestas a cada pregunta.

1. Fallecimientos por COVID-19 en la CDMX (40 % de la tarea).

   1.1. Descarga los datos de la pandemia a nivel nacional: filtra 1) la información de los casos positivos a COVID19 (`CLASIFICACION_FINAL` valores 1, 2, y 3) únicamente para las personas que residen en la CDMX. ¿De cuantos renglones es tu base de datos?

   1.2. Calcula la probabilidad condicional de fallecer por COVID-19, dado cada una de las comorbilidades en la base de datos, i.e. $P(F|C_j)$, $j = 1, 2 \ldots ,$. Obviamente sólo considerando los enfermos de COVID-19 en la CDMX. Haz una gráfica de barras y comenta al respecto.

   1.3. Piensa muy bien la siguiente pregunta, ¿qué comorbilidad esta asociada a un mayor número de fallecimientos por COVID-19 en la CDMX? Justifica tu respuesta de la siguiente manera: sí en la CDMX hubiera 100,000 infectados por COVID-19, obtén el número promedio de fallecimientos asociados a cada comorbilidad.

   El restante 60 % de la tarea

2. En una importante empresa de investigación de biovacunas, es inevitable que los trabajadores estén expuestos a algunos agentes peligrosos, pero altamente tratables, que causan enfermedades. Según los documentos presentados ante las Autoridades de Seguridad y Peligros del estado en el que se encuentra la instalación, el tratamiento proporcionado se adapta a la edad del trabajador (la v.a. X: 0 si es menor de 30 años; 1 si tiene 31 años o más) y la ubicación en la instalación (un sustituto de la virulencia de las cepas patentadas utilizadas en varias partes de la instalación, representado por la v.a. Y = 1, 2, 3 o 4). La composición de los 2,500 empleados en la sede de investigación de la empresa se muestra en la tabla a continuación:

Edad/Localización	1	2	3	4
< 30	150	500	325	250
≥ 31	425	350	300	200

   2.1. Si un trabajador se infecta al azar de manera que el resultado es la variable aleatoria bivariada (X, Y), donde X tiene dos resultados y Y tiene cuatro, obtén la función de densidad conjunta $f(x, y)$ a partir de los datos proporcionados (asumiendo que cada trabajador en cada ubicación tiene igual probabilidad de infección); y determina las funciones de densidad marginales $f(x)$ y $f(y)$.

2.2. ¿Cuál es la probabilidad de que un trabajador que necesita tratamiento haya sido infectado en la ubicación 3 o 4, dado que tiene menos de 30 años?

2.3. Si el costo de tratar a cada trabajador infectado (en dólares por año) se da por la expresión $C = 1500 - 100Y + 500X$, ¿cuánto debería esperar gastar la empresa por trabajador cada año, asumiendo que la composición de los trabajadores permanece igual año tras año?

3. La función de distribución de las variables aleatorias X, Y y Z es:

$$p(1,2,3) = p(2,1,1) = p(2,2,1) = p(2,3,2) = 1/4,$$

Encuentra 2.1) $E[XYZ]$ y 2.2) $E[XY + XZ + YZ]$.

4. La función de distribución conjunta de X y Y está dada por

$$f(x,y) = kxe^{-y}, \quad 0 \leq y \leq x \leq 1,$$

4.1. Encuentra k.

4.2. Encuentra la densidad marginal de X.

4.3. Encuentra la densidad marginal de Y.

4.4. Encuentra $P(Y < 1 | X < 1)$.

4.5. Encuentra $\mathbb{E}(X)$.

4.6. Encuentra $\mathbb{E}(Y)$.

5. El número de reclamaciones recibidas por una compañía de seguros de automóviles en un mes sigue una distribución de Poisson con una media de 20. El 70 % de las pólizas pertenecen al tipo de vehículo A, y el 30 % de las pólizas pertenecen al tipo de vehículo B. Calcula la probabilidad condicional de que más de 10 reclamaciones sean para el tipo de vehículo A dado que al menos 5 de las reclamaciones son para el tipo de vehículo B.

6. Según el Centro Nacional de Estadísticas de Salud de los Estados Unidos, el 25.2 por ciento de los hombres y el 23.6 por ciento de las mujeres nunca desayunan. Supongamos que se eligen muestras aleatorias de 200 hombres y 200 mujeres. Aproximemos la probabilidad de que:

6.1. Al menos 110 de estas 400 personas nunca desayunen;

6.2. El número de mujeres que nunca desayunan sea al menos tan grande como el número de hombres que nunca desayunan.

7. Las variables aleatorias discretas X, Y y Z son independientes si se cumple la siguiente propiedad para todos los valores i, j y k:

$$P(X = i, Y = j, Z = k) = P(X = i)P(Y = j)P(Z = k).$$

Demuestra que si X, Y y Z son independientes, entonces X e Y son independientes, y que esto implica que:

$$P(X = i, Y = j) = P(X = i)P(Y = j).$$

8. Explica cómo generando v.a. $U \sim U(0,1)$ podrías generar v.a. de una distribución de probabilidad dada por:

$$f(y) = 1/(2y + 1), \quad \text{para} \quad 0 < y < 1.$$

9. Un agente de seguros vende una póliza que tiene un deducible de \$100 y un límite de \$5,000. Esto significa que cuando el titular de la póliza presenta un reclamo, el titular de la póliza debe pagar los primeros \$100. Después de los primeros \$100, la compañía de seguros paga el resto del reclamo hasta un pago máximo de \$5,000. Cualquier exceso debe ser pagado por el titular de la póliza. Supongamos que la cantidad en dólares X de un reclamo tiene una distribución continua con una función de distribución de probabilidad

$$f(x) = 1/(1+x)^2, \text{ para } x > 0.$$

Sea Y el monto que la compañía de seguros debe pagar en el reclamo.

9.1. Escribe Y como una función de X, es decir, $Y = h(X)$.

9.2. Encuentra la función de distribución acumulativa de Y.

9.3. Explica por qué Y no tiene una distribución continua ni discreta.

10. Determina la media y varianza de la media muestral $(1/5)\sum_{i=1}^{5} X_i$, donde $X_1, X_2, \ldots, X_5$ son v.a. independientes todas con distribución de probabilidad $f(x) = 4x^3$, con $0 < x < 1$.

11. Sean $X \sim \text{Beta}(\alpha, \beta)$ y $Y \sim \text{Beta}(\alpha + \beta, \eta)$ con X y Y independientes entre sí. Obtén la distribución de la v.a. $W = XY$.