Primer Examen Parcial

Análisis Estadístico Multivariado

Nombre:

Fecha:

Parte 1. Teoría

Ejemplo ilustrativo 1 (cálculo de probabilidades conjuntas)

Sean X, Y variables aleatorias continuas con función de densidad

$$f_{xy}(x, y) = \frac{1}{2}x + \frac{3}{2}y \quad \text{para } 0 \leq x \leq 1, 0 \leq y \leq 1$$

La función de distribución acumulada, o cdf, está dada por

$$F(x, y) = \int_0^x \int_0^y \left(\frac{1}{2}u + \frac{3}{2}v\right) dvdu$$

Para obtener la cdf, integramos primero la parte interna

$$F(x, y) = \int_0^x \left\{\int_0^y \left(\frac{1}{2}u + \frac{3}{2}v\right) dv\right\} du$$

Considerando x constante, la integral nos da

$$F(x, y) = \int_0^x \left\{\frac{1}{2}uv + \frac{3}{2}\frac{v^2}{2}\right\}_0^y du = \int_0^x \left(\frac{1}{2}uy + \frac{3}{4}y^2\right) du$$

La segunda integral nos da

$$F(x,y)=\int_0^x\left(\frac{1}{2}uy+\frac{3}{4}y^2\right)du=\left(\frac{1}{2}\frac{u^2}{2}y+\frac{3}{4}uy^2\right)_0^x=\frac{1}{4}x^2y+\frac{3}{4}xy^2$$

Las funciones marginales se obtienen una variable con respecto a la otra. Para obtener la distribución marginal de x, tenemos la expresión

$$f_x(x)=\int_0^1 f_{xy}(x,y)dy=\int_0^1\left(\frac{1}{2}x+\frac{3}{2}y\right)dy$$

Integrando la expresión, obtenemos

$$f_x(x)=\frac{1}{2}xy+\frac{3}{2}\frac{y^2}{2}\bigg|_0^1=\frac{1}{2}x+\frac{3}{4}$$

Instrucciones

Lea cuidadosamente el problema antes de empezar a resolverlo. Cada problema debe de tener procedimiento completo, adecuado y la respuesta correcta para ser considerado como correcto. No se aceptan dos resultados diferentes para un problema. Escriba claramente y muestre sus operaciones. Escriba sus respuestas en un lugar visible. Marque cada resultado en un recuadro. Todos los resultados finales deben de ser simplificados.

1. (10 puntos) Para la función

$$f_{xy}(x,y)=\frac{1}{2}x+\frac{3}{2}y \quad \text{para } 0\le x\le 1, 0\le y\le 1$$

   a) Encontrar las densidades condicionales de x dado y y de y dado x, que están dadas por

$$f_{x|y}(x\mid y)=\frac{f_{xy}(x,y)}{f_y(y)},\quad f_{y|x}(y\mid x)=\frac{f_{xy}(x,y)}{f_x(x)}$$

   b) Demuestra que X,Y *no* son independientes. Para eso sírvete de la regla que se cunmple cuando son independientes

$$f_{xy}(x,y)=f_x(x)f_y(y)$$

2. (15 puntos)
Un sistema electrónico tiene uno de cada dos tipos diferentes de componentes de operación en operación conjunta. Denote con X y Y las duraciones aleatorias de los componentes del tipo I y tipo II respectivamente. La función de densidad de probabilidad conjunta está dada por

$$f_{xy}(x,y)=\frac{1}{8}x\exp\left[-(x+y)/2\right], \quad \text{para } x,y>0$$

Nota: las mediciones son en cientos de horas.

a) Encontrar $P(X > 1, Y > 1)$. Hint: el límite inferior de ambas integrales es 1.
b) Encontrar la probabilidad de que el componente tipo II tenga una vida útil de más de 200 horas. Hint: se necesita la probabilidad marginal de Y.

3. (15 puntos). La administración en un restaurante de comida rápida está interesada en el comportamiento conjunto de las variables aleatorias X y Y. X, representa el tiempo total entre la llegada de un cliente a la tienda y la salida de la ventanilla de servicio, Y, el tiempo que un cliente espera en la fila antes de llegar a la ventanilla de servicio. Como X incluye el tiempo que un cliente espera en la fila, debemos tener $X \geq Y$. La distribución de valores observados puede ser modelada por la siguiente función de densidad de probabilidad conjunta

$$f_{xy}(x, y) = \exp(-x), \quad 0 \leq y \leq x < \infty$$

con el tiempo medido en minutos.

a) $P(X < 2, Y > 1)$.
b) Las funciones de densidad marginal para ambas variables.
c) La función de densidad condicional de X dado $Y = y$.

4. (15 puntos). La densidad conjunta para las variables aleatorias X y Y, donde X es el cambio unitario de temperatura y Y es la proporción de desplazamiento espectral que produce cierta partícula atómica está dada por

$$f_{XY}(x, y) = 10xy^2, \ 0 < x < y < 1$$

Dibuje la región $0 < x < y < 1$ en el plano x, y y calcule las siguientes probabilidades, y dé una interpretación de los resultados obtenidos
a) $P(X \leq 1/2, Y \leq 1/2)$
b) $P(1/4 \leq Y \leq 3/4)$
c) $P(X \leq 3/4)$

5. (15 puntos). Un supermercado cuenta tanto con una caja normal como con una rápida. La función de masa de probabilidad conjunta de X_1 el número de clientes formados en la caja normal a una hora particular del día y de X_2 el número de clientes formados en la caja rápida a la misma hora está dada por

$X_1 \backslash X_2$	0	1	2	3
0	0.08	0.07	0.04	0
1	0.06	0.15	0.05	0.04
2	0.05	0.04	0.10	0.06
3	0	0.03	0.04	0.07
4	0	0.01	0.05	0.06

En donde X_1 son las filas y X_2 las columnas.
a) Calcular $P(X_1 = 1, X_2 = 1)$

b) Calcular $P(X_1 \leq 3, X_2 \leq 2)$
c) Determinar la función de probabilidad marginal de X_1 y X_2
d) Determinar la función de probabilidad condicional para $X_2 \mid X_1 = 2$. Dar una interpretación al resultado.
e) ¿Cuál es la probabilidad de que dos clientes estén formados en la caja rápida *dado* que tres clientes estén formados en la caja normal?
g) ¿X_1 y X_2 son independientes? Justificar la respuesta.

6. (10 puntos). En la figura 1 se muestra una muestra bivariada con $\boldsymbol{\mu} = (1,1), \rho = -0.6, \sigma_1 = 2$ y $\sigma_2 = 8$. Los puntos azul y naranja tienen coordenadas $p_1 =$(4, 15) y $p_2 =$(4, -13) respectivamente.

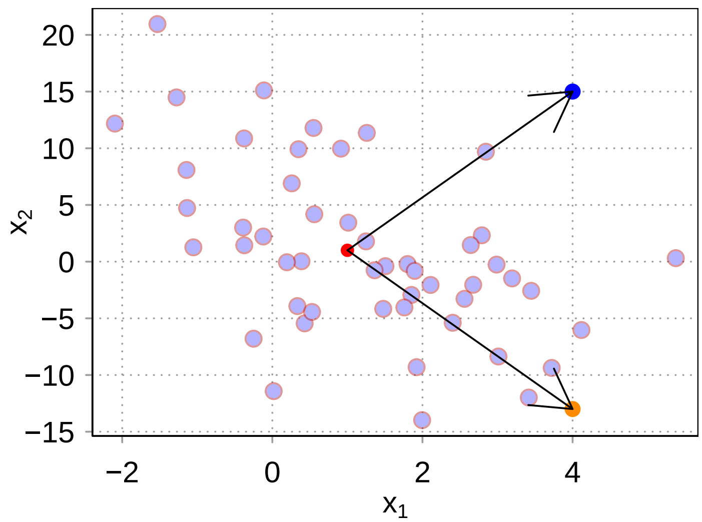

Figure 1: Muestra bivariada $X_1, X_2 \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$

a) Encuentra las distancias euclideanas (norma L_2) y de Mahalanobis d para cada punto.
b) Concluye si, con un $\alpha = 0.01$, se debería clasificar alguno de los puntos como un *outlier*. (Para esto, recuerda que d^2 tiene distribución $\chi^2(1-\alpha, df)$).

Parte 2. Práctica

Emotions sensor data set

Context

Emotions detection is an interesting blend of psychology and technology.

This technology helps to build a companion robots, this robots can be friendly and have the ability to recognize users' emotions and needs, and to act accordingly.

It's essentially a way to determine how your consumers are reacting to your website, social media posts, and other forms of your online content this helps to transform the face of marketing and advertising by reading human emotions and then adapting consumer experiences to these emotions in real time.

What is Emotions Sensor Dataset:

Emotions Sensor Data Set contain top 23 730 english words classified statistically using naive Bayes algorithm into 7 basic emotion disgust, surprise, neutral, anger, sad, happy and fear.

How We Build This Dataset:

First we collected thousands of sentences, blogs and twitters,all about 1,185,540 words. We labeled manually and automatically this sentences Into 7 basic emotion disgust, surprise, neutral ,anger, sad, happy and fear. Now we choose the top of most used 23 730 english words Word by word we calculated the probabilities of existence of this words in disgust, surprise, neutral, anger, sad, happy and fear sentences and put them in simple CSV file.

We used the Naive Bayes Algorithm to calculate this probabilities of existence of this words.

Ejercicio (20 puntos)

Para la base de datos de `Andbrain_DataSet.csv`

a) Calcule las medidas de tendencia central (todas las variables) y con base en la descripción *emotions sensor data set* dé una interpretación de los resultados obtenidos para una de las variables.

b) ¿Para qué sirven las medidas de tendencia central?

c) Calcule las medidas de dispersión univariantes. ¿Cómo interpreta los valores obtenidos de la varianza y la desviación estándar?

d) ¿Para qué sirven las medidas de dispersión?

e) Calcule las medidas de dispersión bivariantes (matriz de covarianza, matriz de correlación, vector de medias).

f) ¿Qué variables presentan una relación directa y qué variables presentan relación inversa?

g) ¿Qué variables tienen una relación débil?

h) ¿Qué variables tienen una relación fuerte?

i) De una interpretación escrita de los resultados obtenidos en la matriz de correlación.

j) ¿Cuál es la principal diferencia entre covarianza y correlación?

k) ¿Qué valores puede tomar la covarianza y qué valores puede tomar la correlación?

l) Si dos variables tienen correlación cero, ¿se puede concluir que son independientes? Justificar