Análisis Estadístico Multivariante

Francisco Javier Mercader Martínez

Índice

Tema 1: Vectores aleatorios

Un vector aleatorio (v.a.) k-dimensional sobre un espacio de probabilidad $(\Omega, \mathcal{S}, \mathcal{P})$ es $X = (X_1, \ldots, X_k)$ tal que

$$X_i^{-1}(-\infty, x] \in \mathcal{S}$$

para todo $x \in \mathbb{R}$, $i = 1, \ldots, k$

- Función de distribución conjunta

$F : \mathbb{R}^k \longrightarrow [0,1]$

$$F(x_1, \ldots, x_k) \coloneqq P[X_]$$

1.1) Independencia de las variables aleatorias

- Definición

Las variables aleatorias $X_1, \ldots, X_k$ son independientes si los sucesos

$$\{x_1 \leq x_1\}, \{X_2 \leq x_2\}, \ldots, \{X_k \leq x_k\}$$

son idependientes para todo $x_1, \ldots, x_k \in \mathbb{R}$.

Esto es equivaletne a que

$$F(x_1, \ldots, x_k) = P[X_1 \leq x_1] \cdot P[X_2 \leq x_2] \cdots P[X_k \leq x_k]$$

para todo $x_1, \ldots, x_k \in \mathbb{R}$.

- Distribuciones marginales

La función $F_{X_i}(x_i) = P[X_i \leq x_i]$ se denomina función de distribución marginal i-ésima y corresponde con la función de distribución de la variable aleatoria X_i

Las distribuciones marginales pueden obtenerse a partir de la distribución conjunta:

$$F_{X_i}(x_I) = F(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty)$$

Análogamente, la función de distribución marginal del subvector aleatorio $(X_{i_1}, \ldots, X_{i_m})$ vendrá dada por

$$F_{x_{i_1}, \ldots, X_{i_m}}$$

1.2) Vector aleatorio absolutamente continuo

Un vector aleatorio X es absolutamente continuo si existe una función $f : \mathbb{R}^k \longrightarrow \mathbb{R}$ no negativa (llamada función de densidad) tal que

$$F(x) = F(x_1, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(z_1, \ldots, z_k) \mathrm{d}z_k, \ldots, \mathrm{d}z_1,$$

para todo $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$

Usando el teorema fundamental del cálculo, se tiene que en cada punto de continuidad $(x_1, \ldots, x_k)$ de f:

$$\frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1, \ldots, \partial x_k}$$

Existen variables aleatorias cuya función de distribución es continua pero que no son absolutamente continuas

1.3) Vector aleatorio discreto

Un vector aleatorio X se dice que es discreto si existe un conjunto numerable $\mathcal{S} \in \mathbb{R}^k$ tal que $P(X \in \mathcal{S}) = 1$.

Función masa de probabilidad de una vector aleatorio discreto:

$$P[X = x] = P[X_1 = x_1, \dots, X_k = x_k]$$

para todo $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, satisfaciendo:

→ $P[X = x] \geq 0, \ \forall x \in \mathcal{S}$

→ $\sum_{x \in \mathcal{S}} P[X = x] = 1$

Función de distribución de un vector aleatorio discreto:

$$F(x) = P[X \leq x] =$$

1.4) Distribuciones marginales

1.4.1) Caso continuo

- Distribución marginal de la variable aleatoria X_i

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio continuo con función de densidad f entonces cada componente X_i es de tipo continuo y su función de distribución es;

$$F_{X_i}(x_i) = P[X_i \leq x_i] = \int_{-\infty}^{x_i} f_{X_i}(z_i) \mathrm{d}z_i,$$

con

$$f_{X_i} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(z_1, \dots, z_k) \mathrm{d}z_1, \dots,$$

Nota:
A y B independientes $\longleftrightarrow P(A|B) = P(A)$
$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = P(A) \longrightarrow P(A \cap B) = P(A) \cdot P(B)$

- Distribución condicionada al valor de una variable

Sea $X = (X_1, \dots, X_k)$

- Distribución condicionada a valores de varias variables

Sea $X = (X_1, \dots, X_k)$ un vecotr aleatorio continuo

RELACIÓN DE PROBLEMAS: VECTORES ALEATORIOS
ANÁLISIS ESTADÍSTICO MULTIVARIANTE
GRADO EN CIENCIA E INGENIERÍA DE DATOS

1. Sea (X, Y) un vector aleatorio con función de densidad conjunta
$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$
Hallar las distribuciones marginales y condicionadas.

2. Obtener las distribuciones marginales y condicionadas asociadas al vector aleatorio (X, Y) con función de densidad
$$f(x,y) = \begin{cases} 2 & \text{si } 0 < x < 1, 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

3. Sea (X, Y) un vector aleatorio con función de densidad
$$f(x,y) = \begin{cases} \frac{3}{4}\left[xy + \frac{x^2}{2}\right] & \text{si } 0 < x < 1, 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$
Hallar la distribución marginal de X y la distribución de Y condicionada a $X = \frac{1}{2}$.

4. Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio con función masa de probabilidad
$$P[X_1 = x_1, X_2 = x_2] = \frac{k}{2^{x_1+x_2}}, x_1, x_2 \in \mathbb{N},$$
donde k es una constante. Obtener las distribuciones marginales y condicionadas.

5. Calcular la función de densidad de una distribución normal bidimensional en (1, 1) si las medias son cero, las varianzas 1 y 4, y la covarianza 1.

6. Sea (X, Y) un vector aleatorio con distribución uniforme en el cuadrado unidad, $[0, 1] \times [0, 1]$, con función de densidad conjunta
$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$
Calcular el valor esperado de $g(X, Y) = XY^2$, es decir, $E[XY^2]$.

7. (X, Y) vector aleatorio discreto con función masa de probabilidad conjunta:

$X\backslash Y$	1	2
1	1/9	2/9
2	2/9	4/9

   a) Calcular $E[X + Y]$, $E[2X + 3Y]$.
   b) Obtener el vector de medias, la matriz de covarianzas y la matriz de correlaciones del vector (X, Y).
   c) ¿Son independientes? ¿Están incorreladas?

8. Demostrar que el vector de medias muestral es el punto de $\mathbb{R}^k$ que minimiza la suma de las distancias al cuadrado (error cuadrático medio, MSE).

1) Sea (X, Y) un vector aleatorio con función de densidad conjunta

$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1, \ 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Hallar las distribuciones marginales y condicionadas

$$\underset{0<x<1}{f_X(x)} = \int_{\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \cancel{\int_{-\infty}^{0} 0\,\mathrm{d}y} + \int_0^1 1\,\mathrm{d}y + \cancel{\int_{0}^{+\infty} 0\,\mathrm{d}y} = [y]_{y=0}^{y=1} = 1 \qquad f_X(x) = \begin{cases} 1 & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{0<y<1}{f_Y(y)} = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x = \int_0^1 1\,\mathrm{d}x = [x]_{x=0}^{x=1} = 1 \qquad f_Y(y) = \begin{cases} 1 & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$f(x,y) = \begin{cases} 1 & 0 < x < 1, \ 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{f_X(x^*)>0}{y|x = x^*} \longrightarrow \underset{0<x^*<1}{f_{y|x=x^*}} = \frac{f(x^*,y)}{f_X(x^*)} = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$f(x,y) = f_X(x) \cdot f_Y(y)$ $\quad X$ e Y independientes

2) Obtener las distribuciones marginales y condicionadas asociadas al vector aleatorio (X, Y) con función de densidad

$$f(x,y) = \begin{cases} 2 & \text{si } 0 < x < 1, \ 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{0<x<1}{f_X(x)} = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \int_0^x 2\,\mathrm{d}y = [2y]_{y=0}^{y=x} = 2x \longrightarrow \begin{cases} 2x & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{0<y<1}{f_Y(y)} = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x = \int_y^1 2\,\mathrm{d}x = [2x]_{x=y}^{x=1} = 2 - 2y \longrightarrow \begin{cases} 2-2y & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Los recintos son dependientes.

$y|x = x^*$

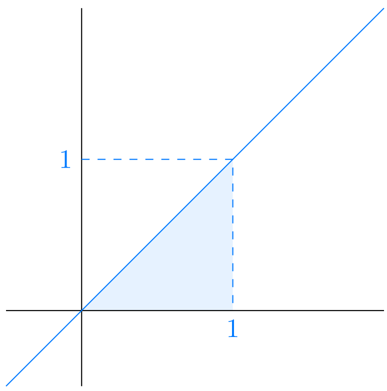
$f_X(x^*) > 0$

$$\underset{0<x^*<1}{f_{y|x=x^*}(y|x^*)} = \frac{f(x^*,y)}{f(x^*)} = \begin{cases} \dfrac{2}{2x^*} & 0 < y < x^* \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \dfrac{1}{x^*} & 0 < y < x^* \\ 0 & \text{en otro caso} \end{cases}$$

$x|y = y^*$

$f_Y(y^*) > 0$

$$\underset{0<y^*<1}{f_{x|y=y^*}(x|y^*)} = \frac{f(x,y^*)}{f_Y(y^*)} = \begin{cases} \dfrac{2}{2-2y^*} & \text{si } y^* < x < 1 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \dfrac{1}{1-y^*} & \text{si } y^* < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$



3) Sea (X,Y) un vector aleatorio con función de densidad

$$f(x,y) = \begin{cases} \frac{3}{4}\left[xy + \frac{x^2}{2}\right] & \text{si } 0<x<1,\ 0<y<2 \\ 0 & \text{en otro caso} \end{cases}$$

Hallar la distribución marginal de X y la distribución de Y condicionada a $X = \frac{1}{2}$.

$$\underset{0<x<1}{f_X(x)} = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \int_0^2 \frac{3}{4}\left[xy + \frac{x^2}{2}\right] = \frac{3}{4}\left[\frac{xy^2}{2} + \frac{x^2}{2}\right]_{y=0}^{y=2} = \frac{3}{4}\left(2x + \frac{x^2}{2}\right) \longrightarrow \begin{cases} \frac{3}{4}\left(2x + \frac{x^2}{2}\right) & \text{si } 0<x<1 \\ 0 & \text{en otro caso} \end{cases}$$

$y|x = x^*$

$$\underset{0<x^*<1}{f_{y|x=x^*}(y|x^*)} = \frac{f(x^*,y)}{f(x^*)} = \frac{\frac{3}{4}\left(x^*y + \frac{(x^*)^2}{2}\right)}{\frac{3}{4}\left(2x^* + \frac{(x^*)^2}{2}\right)} = \frac{x^*y + \frac{(x^*)^2}{2}}{2x^* + \frac{(x^*)^2}{2}} = \frac{2x^*y + (x^*)^2}{4x^* + (x^*)^2} \xrightarrow{x^*=\frac{1}{2}} \frac{y + \frac{1}{4}}{2 + \frac{1}{4}} = \frac{4y+1}{9}$$

$$\begin{cases} \frac{4y+1}{9} & \text{si } 0<y<2 \\ 0 & \text{en otro caso} \end{cases}$$

4) Sea $X = (X_1, X_2)$ un vector aleatorio con función masa de probabilidad

$$P[X_1 = x_1, X_2 = x_2] = \frac{k}{2^{x_1+x_2}}, x_1, x_2 \in \mathbb{N}$$

donde k es una constante. Obtener las distribuciones marginales y condicionadas.