Universidad Nacional de Colombia
Departamento de Estadística

DIPLOMADO EN PROBABILIDAD E INFERENCIA BÁSICA

Módulo

PROBABILIDAD

Quinta sesión

Vectores aleatorios bidimensionales

Generalidades

En ocasiones, en el mismo conjunto de unidades físicas se mide u observa más de una única característica. En este caso, el análisis se centra en las posibles asociaciones entre variables.

Vector aleatorio p −dimensional

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad en el cual se han definido las variables aleatorias $X_1, \dots\ X_p$, con $p \in \mathbb{Z}^+$. El objeto $\mathbf{X} = \begin{pmatrix} X_1 & \dots & X_p \end{pmatrix}^T$ se llama *vector aleatorio de dimensión p*. El *rango* del vector $\mathbf{X}$ es

$$R_{\mathbf{X}} = \{(x_1, \dots, x_k) : x_i \in R_i, i = 1, \dots, p\}$$

donde $x_i \in \mathbb{R}$ y R_i es el rango de la variable X_i.

Al igual que en el caso unidimensional, es claro que todo vector aleatorio $\mathbf{X}$ de dimensión p tiene asociado un espacio de probabilidad compuesto por el rango del vector $R_{\mathbf{X}} \subseteq \mathbb{R}^{\boldsymbol{p}}$, la σ-álgebra de eventos sobre $R_{\mathbf{X}}$ (formada por hipercubos de dimensión p) y la medida de probabilidad $P_{\mathbf{X}}$ inducida por el vector aleatorio.

Vector aleatorio bidimensional

En el caso para el cual $p = 2$ el vector aleatorio es *bidimensional* y su rango es un subconjunto del plano euclidiano $\mathbb{R} \times \mathbb{R}$.

Si el conjunto de valores posibles de $\mathbf{X} = (X_1, X_2)$ es finito o contable, entonces el vector será *discreto*. En este caso, los valores del vector $\mathbf{X}$ son $\left(x_{1_i}, x_{2_j}\right)$, con $i = 1, 2, \ldots$ y $j = 1, 2, \ldots$.

Si los valores posibles del vector $\mathbf{X} = (X_1, X_2)$ forman algún conjunto innumerable del plano euclidiano $\mathbb{R} \times \mathbb{R}$, entonces el vector aleatorio es *continuo*. En este caso, a manera de ilustración, si $a \leq x_1 \leq b$ y $c \leq x_2 \leq d$, entonces el rango del vector está constituido por todos los pares de números reales (x_1, x_2) en el plano $[a; b] \times [c, d]$.

Distribución conjunta

La *distribución conjunta* muestra el comportamiento probabilístico simultáneo de las variables que integran un vector aleatorio

Distribución de probabilidad conjunta

Sea el vector aleatorio discreto $\mathbf{X}$ =(X_1, X_2). Si cada valor posible de $\mathbf{X}$ se asocia con un número

$$p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) = P\left(X_1 = x_{1_i};\ X_2 = x_{2_j}\right)$$

tal que

- ❑ $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) \geq 0$ para todo i, j
- ❑ $\sum_i \sum_j p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) = 1$

entonces el arreglo tabular $\left[\left(x_{1_i}, x_{2_j}\right);\ p_{ij}\right]$, para todo i, j, forman la *distribución de probabilidad conjunta* del vector aleatorio discreto $\mathbf{X}$.

Ejemplo

<u>EJEMPLO 1 (E1)</u>. Se seleccionan aleatoriamente tres balotas de una urna que contiene tres rojas, cuatro blancas y cinco azules. Sea X la cantidad de balotas rojas y Y la cantidad de balotas blancas seleccionadas de la urna.

Si se denota con $p(x, y)$ la probabilidad de seleccionar x balotas blancas y y balotas rojas de la urna, entonces la función de probabilidad conjunta de X y Y está dada por

$$p(x, y) = \begin{cases} \dfrac{C_x^3 C_y^4 C_{3-x-y}^5}{C_3^{12}} & x, y \geq 0, x + y = 3 \\ 0 & \text{en otro caso} \end{cases}$$

Continuación del ejemplo

$p_{X,Y}(x,y)$		Y			
		0	1	2	3
X	0	$1/22$	$3/22$	$3/44$	$1/220$
	1	$2/11$	$3/11$	$3/55$	0
	2	$9/110$	$3/22$	0	0
	3	$1/55$	0	0	0

Función de densidad conjunta

El vector aleatorio X $=(X_1, X_2)$ es conjuntamente continuo con rango $R_{\mathbf{X}} \in \mathbb{R} \times \mathbb{R}$ en el plano euclidiano si existe una función $f_{\mathbf{X}}(x_1, x_2) \geq 0$, definida para todo $(x_1, x_2) \in R_{\mathbf{X}}$, tal que

$$P(\mathbf{X} \in \mathcal{R}) = \iint_{\mathcal{R}} f_{\mathbf{X}}(x_1, x_2)\, dx_1 dx_2 \qquad \iint_{R_{\mathbf{X}}} f_{\mathbf{X}}(x_1, x_2)\, dx_1 dx_2 = 1$$

La función $f_{\mathbf{X}}(x_1, x_2)$ se llama *función de densidad conjunta* del vector aleatorio bidimensional $\boldsymbol{X} = (X_1, X_2)$.

Ejemplo

EJEMPLO 2 (E2). Supóngase que las variables aleatorias X y Y tienen función de densidad conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{2}y^2 & 0 \leq x \leq 2; 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Luego,

$$P\left(0 \leq X \leq 1; 0 \leq Y \leq \frac{1}{2}\right) = \frac{3}{2}\int_0^1 \int_0^{\frac{1}{2}} y^2 dydx = \frac{1}{16}$$

Distribuciones marginales

Se conocen como *distribuciones marginales* a las distribuciones de cada una de las variables aleatorias que integran un vector

Concepto de distribución marginal

CASO DISCRETO

Sea el vector aleatorio discreto $\mathbf{X}=(X_1, X_2)$ con distribución conjunta dada por $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$. Las distribuciones marginales de X_1 y X_2, respectivamente, son

$$p_{X_1}\left(x_{1_i}\right) = \sum_j p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) \quad i = 1, 2, \ldots$$

$$p_{X_2}\left(x_{2_j}\right) = \sum_i p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) \quad j = 1, 2, \ldots.$$

CASO CONTINUO

Sea el vector aleatorio continuo $\mathbf{X}=(X_1, X_2)$ con función de densidad conjunta dada por $f_{\mathbf{X}}(x_1, x_2)$. Las distribuciones marginales de X_1 y X_2, respectivamente, son

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2)\,dx_2$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2)\,dx_1$$

Ejemplo para el caso discreto

En el ejemplo **E1,** las funciones de masa de probabilidad marginales, respectivamente, son

$$p_X(x) = \begin{cases} \dfrac{C_x^3 C_{3-x}^9}{C_3^{12}} & x = 0,1,2,3 \\ 0 & \text{en otro caso} \end{cases}$$

$$p_Y(y) = \begin{cases} \dfrac{C_y^4 C_{4-y}^9}{C_3^{12}} & y = 0,1,2,3 \\ 0 & \text{en otro caso} \end{cases}$$

$p_{X,Y}(x,y)$		Y				$p_X(x)$
		0	**1**	**2**	**3**	
X	**0**	$^1/_{22}$	$^3/_{22}$	$^3/_{44}$	$^1/_{220}$	$^{14}/_{55}$
	1	$^2/_{11}$	$^3/_{11}$	$^3/_{55}$	0	$^{28}/_{55}$
	2	$^9/_{110}$	$^3/_{22}$	0	0	$^{12}/_{55}$
	3	$^1/_{55}$	0	0	0	$^1/_{55}$
$p_Y(y)$		$^{36}/_{110}$	$^6/_{11}$	$^{27}/_{220}$	$^1/_{220}$	1

Ejemplo para el caso continuo

En el ejemplo **E2,** se definió el vector aleatorio continuo (X, Y) con función de densidad conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{2}y^2 & 0 \le x \le 2; 0 \le y \le 1 \\ 0 & \text{en otro caso} \end{cases}$$

La función de densidad marginal de X es

$$f_X(x) = \frac{3}{2}\int_0^1 y^2 dy = \frac{1}{2}, \qquad 0 \le x \le 2$$

y es cero en otro caso.

La función de densidad marginal de Y es

$$f_Y(y) = \frac{3}{2}y^2 \int_0^2 dx = 3y^2, \qquad 0 \le y \le 1$$

y es cero en otro caso.

Distribuciones condicionales

Cuando se trabaja con dos variables aleatorias distribuidas conjuntamente, puede ser de interés encontrar cómo cambia la distribución de una de ellas dado un valor particular de la otra

Distribuciones condicionales discretas

Sea el vector aleatorio discreto $\mathbf{X} = (X_1, X_2)$ con distribución conjunta dada por $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$. Las distribuciones condicionales son

$$p_{X_2|X_1=x_{1_i}}\left(x_{2_j}\right) = \frac{p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)}{p_{X_1}\left(x_{1_i}\right)} \quad i = 1, 2, \dots ; j = 1, 2, \dots$$

$$p_{X_1|X_2=x_{2_j}}\left(x_{1_i}\right) = \frac{p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)}{p_{X_2}\left(x_{2_j}\right)} \quad i = 1, 2, \dots ; j = 1, 2, \dots$$

donde $p_{X_1}\left(x_{1_i}\right) > 0$ y $p_{X_2}\left(x_{2_j}\right) > 0$.

Ejemplo

En el ejemplo **E1**, de la distribución marginal de X se tiene que $P(X = 1) = p_X(1) = \frac{28}{55}$

Luego, la función de masa de probabilidad condicional de Y dado $X = 1$ está dada por

$$p_{Y|X=1}(y = 0) = \frac{p_{X,Y}(1,0)}{p_X(1)} = \frac{P(X = 1; Y = 0)}{P(X = 1)} = \frac{\frac{2}{11}}{\frac{28}{55}} = \frac{5}{14}$$

$$p_{Y|X=1}(y = 1) = \frac{p_{X,Y}(1,1)}{p_X(1)} = \frac{P(X = 1; Y = 1)}{P(X = 1)} = \frac{\frac{3}{11}}{\frac{28}{55}} = \frac{15}{28}$$

$$p_{Y|X=1}(y = 2) = \frac{p_{X,Y}(1,2)}{p_X(1)} = \frac{P(X = 1; Y = 2)}{P(X = 1)} = \frac{\frac{3}{55}}{\frac{28}{55}} = \frac{3}{28}$$

Distribuciones condicionales continuas

Sea el vector aleatorio continuo $\mathbf{X}$ =(X_1, X_2) con función de densidad conjunta dada por $f_{\mathbf{X}}(x_1, x_2)$. Las distribuciones condicionales son

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_1}(x_1)}$$

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_2}(x_2)}$$

donde $f_{X_1}(x_1) > 0$ y $f_{X_2}(x_2) > 0$.

Ejemplo

La función de densidad conjunta de las variables X y Y está dada por

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}x(2-x-y) & 0 \le x \le 1; 0 \le y \le 1 \\ 0 & \text{en otro caso} \end{cases}$$

Luego, la función de densidad marginal de Y es

$$f_Y(y) = \frac{12}{5}\int_0^1 x(2-x-y)dx = \frac{2}{5}(4-3y)$$

Continuación del ejemplo

La distribución condicional de X dado Y es

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{x(2-x-y)}{6(4-3y)}$$

Para un valor particular de la variable Y, digamos $Y = 1$, la función de densidad condicional de $X|Y = 1$ es una función de los valores x; así, se tiene que

$$f_{X|Y=1}(x) = \frac{x}{6}(1-x), \qquad 0 \leq x \leq 1$$

Independencia de variables aleatorias

Dos variables aleatorias pueden exhibir comportamientos probabilísticos *independientes* si la distribución de una de ellas se mantiene invariante cuando la otra asume un valor particular. En caso contrario, se dice que las variables están correlacionadas o, simplemente, asociadas.

Concepto de independencia

CASO DISCRETO

Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio discreto con función de probabilidad conjunta dada por $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$ y funciones de masa marginales dadas, respectivamente, por $p_{X_1}\left(x_{1_i}\right)$ y $p_{X_2}\left(x_{2_j}\right)$. Las variables X_1 y X_2 son independientes, si y solo si se cumple que

$$p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) = p_{X_1}\left(x_{1_i}\right) p_{X_2}\left(x_{2_j}\right)$$

para todo $(x_1, x_2) \in R_{X_1} \times R_{X_2}$.

CASO CONTINUO

Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio continuo con función de densidad conjunta dada por $f_{\mathbf{X}}(x_1, x_2)$ y funciones de densidad marginales dadas, respectivamente, por $f_{X_1}(x_1)$ y $f_{X_2}(x_2)$. Las variables X_1 y X_2 son independientes, si y solo si se cumple que

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

para todo $(x_1, x_2) \in R_{X_1} \times R_{X_2}$.

Ejemplo para el caso discreto

Sean X y Y variables aleatorias discretas con función de probabilidad conjunta

$$p_{X,Y}(x,y) = \begin{cases} \frac{1}{30}(x+y) & x = 0,1,2; y = 0,1,2,3 \\ 0 & \text{en otro caso} \end{cases}$$

De acuerdo con la distribución marginal de X se tiene que

$$p_X(0) = \sum_{y=0}^{3} p_{X,Y}(0,y) = \frac{1}{5}$$

De acuerdo con la distribución marginal de Y se tiene que

$$p_Y(0) = \sum_{x=0}^{2} p_{X,Y}(x,0) = \frac{1}{10}$$

Luego, X y Y no son independientes puesto que $p_{X,Y}(0,0) \neq p_X(0)p_{X_2}(0)$.

Ejemplo para el caso continuo

Sean X y Y variables aleatorias continuas con función de densidad conjunta

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{2}y^2 & 0 \leq x \leq 2; 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

La distribución marginal de X es

$$f_X(x) = \frac{3}{2}\int_0^1 y^2 dy = \frac{1}{2}y^3 \Big|_0^1 = \frac{1}{2}$$

La distribución marginal de Y es

$$f_Y(y) = \frac{3}{2}y^2 \int_0^2 dx = \frac{3}{2}y^2 x \Big|_0^2 = 3y^2$$

Luego, es evidente que $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ para todo $(x,y) \in [0;2] \times [0;1]$.

Consecuencias de la independencia

Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio discreto con función de probabilidad conjunta dada por $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$ y funciones de masa marginales dadas, respectivamente, por $p_{X_1}\left(x_{1_i}\right)$ y $p_{X_2}\left(x_{2_j}\right)$. Las variables X_1 y X_2 son independientes, si y solo si se cumple que

$$p_{X_1|X_2=x_{2_j}}\left(x_{1_i}\right) = p_{X_1}\left(x_{1_i}\right)$$

$$p_{X_2|X_1=x_{1_i}}\left(x_{2_j}\right) = p_{X_2}\left(x_{2_j}\right)$$

para todo i y j

Consecuencias de la independencia

De igual modo, sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio continuo con función de densidad conjunta dada por $f_{\mathbf{X}}(x_1, x_2)$ y funciones de densidad marginales dadas, respectivamente, por $f_{X_1}(x_1)$ y $f_{X_2}(x_2)$. Las variables X_1 y X_2 son independientes, si y solo si se cumple que

$$f_{X_1|X_2=x_2}(x_1) = f_{X_1}(x_1)$$

$$f_{X_2|X_1=x_1}(x_2) = f_{X_2}(x_2)$$

para todo x_1 y x_2

Medidas de asociación

Valor esperado de una función

Sean $\mathbf{X} = (X_1, X_2)$ un vector aleatorio y $g(\mathbf{X})$ una función de valor real que es una variable aleatoria. Si $\mathbf{X}$ es un vector discreto con función de distribución conjunta $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$, se define el valor esperado de la función $g(\mathbf{X})$ como:

$$E[g(\mathbf{X})] = \sum_i \sum_j g\left(x_{1_i}, x_{2_j}\right) p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$$

Y si $\mathbf{X}$ es un vector continuo con función de densidad conjunta $f_{\mathbf{X}}(x_1, x_2)$, entonces

$$E[g(\mathbf{X})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f_{\mathbf{X}}(x_1, x_2) dx_1 dx_2$$

Caso particular

Sean $\mathbf{X} = (X_1, X_2)$ un vector aleatorio y $g(\mathbf{X}) = X_1$. Si $\mathbf{X}$ es un vector discreto con función de distribución conjunta $p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right)$, entonces

$$E[g(\mathbf{X})] = \sum_{i} \sum_{j} x_{1_i}\, p_{\mathbf{X}}\left(x_{1_i}, x_{2_j}\right) = E(X_1)$$

Si $\mathbf{X}$ es un vector continuo con función de densidad conjunta $f_{\mathbf{X}}(x_1, x_2)$, entonces

$$E[g(\mathbf{X})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f_{\mathbf{X}}(x_1, x_2)\, dx_1 dx_2 = E(X_1)$$

Coeficiente de correlación lineal

Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio. Se define el *coeficiente de correlación lineal* entre las variables X_1 y X_2 como

$$\rho_{12} = \frac{E(X_1X_2) - E(X_1)E(X_2)}{\sqrt{V(X_1)}\sqrt{V(X_2)}}$$

El coeficiente de correlación lineal mide el grado de intensidad de la relación lineal que existe entre las variables X_1 y X_2, pero nunca establece relaciones de causalidad.

El coeficiente de correlación lineal entre las variables X_1 y X_2. Se cumple que

$$-1 \leq \rho_{12} \leq 1$$

Si $X_2 = aX_1 + b$, entonces $|\rho_{12}| = 1$. En este caso, si $a > 0$, entonces $\rho_{12} = 1$ y si $a < 0$, entonces $\rho_{12} = -1$.

Además, si X_1 y X_2 son variables aleatorias independientes, entonces $\rho_{12} = 0$.

Ejemplo

Sean X y Y variables aleatorias continuas con función de densidad conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} \frac{8}{3}xy & 0 \leq x \leq 1; x \leq y \leq 2x \\ 0 & \text{en otro caso} \end{cases}$$

Se puede demostrar que $\rho_{XY} = 0{,}7393$. Luego, entre la variables X y Y variables existe una asociación proporcional directa de tipo lineal.

GRACIAS

Diagramado en

PowerPoint