Variables bidimensionales discretas

Sean X e Y dos variables aleatorias discretas. Su comportamiento puede ser descripto por la **función de probabilidad conjunta**

$$p_{X,Y}(x,y) = P\big(X = x, Y = y\big) \qquad x \in \mathcal{R}_X,\, y \in \mathcal{R}_Y. \tag{1}$$

La funciones de probabilidad **marginales** pueden ser obtenidas como

$$p_X(x) = P\big(X = x\big) = \sum_{y \in \mathcal{R}_Y} p_{X,Y}(x,y) \qquad x \in \mathcal{R}_X, \tag{2}$$

$$p_Y(y) = P\big(Y = y\big) = \sum_{x \in \mathcal{R}_X} p_{X,Y}(x,y) \qquad y \in \mathcal{R}_Y. \tag{3}$$

Las variables aleatorias son **independientes** sii

$$p_{X,Y}(x,y) = p_X(x) p_Y(y) \qquad x \in \mathcal{R}_X,\, y \in \mathcal{R}_Y. \tag{4}$$

Es fácil ver que, en este caso,

$$P\big((X,Y) \in [a,b] \times \in [c,d]\big) = P\big(X \in [a,b], Y \in [c,d]\big) = P\big(X \in [a,b]\big) \cdot P\big(Y \in [c,d]\big),$$

para todo par de intervalos $[a,b], [c,d] \subset \mathbb{R}$.
El valor esperado de una función h de ambas variables aleatorias ($h : \mathbb{R}^2 \to \mathbb{R}$) viene dado por

$$E\left[h(X,Y)\right] = \sum_{(x,y) \in \mathcal{R}_X \times \mathcal{R}_Y} h(x,y) p_{X,Y}(x,y) = \sum_{x \in \mathcal{R}_X} \sum_{y \in \mathcal{R}_Y} h(x,y) p_{X,Y}(x,y). \tag{5}$$

Sean dos funciones $k, m : \mathbb{R} \to \mathbb{R}$. Luego, si X e Y son variables aleatorias independientes,

$$E\left[k(X) m(Y)\right] = \sum_{x \in \mathcal{R}_X} \sum_{y \in \mathcal{R}_Y} k(X) m(Y) p_X(x) p_Y(y) = E\left[k(X)\right] E\left[m(Y)\right]. \tag{6}$$

Variables bidimensionales continuas

El comportamiento de dos variables aleatorias continuas X e Y suele se descripto por una **densidad de probabilidad conjunta** $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}^{\geq 0}$. A partir de ésta, pueden calcularse todas las probabilidades relacionadas a ambas variables como

$$P\big((X,Y) \in B\big) = \iint\limits_{(x,y) \in B} f_{X,Y}(x,y)\, dx dy \qquad B \subseteq \mathbb{R}^2. \tag{7}$$

Allí donde las derivadas tienen sentido, la densidad conjunta puede escribirse como

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} P\big(X \leq x, Y \leq y\big). \tag{8}$$

La densidades de probabilidad **marginales** pueden ser obtenidas como

$$f_X(x) = \int\limits_{\mathbb{R}} f_{X,Y}(x,y)\, dy, \qquad f_Y(y) = \int\limits_{\mathbb{R}} f_{X,Y}(x,y)\, dx. \tag{9}$$

Variables bidimensionales continuas (continuado)

Las variables aleatorias son **independientes** sii

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \qquad (x,y) \in \mathbb{R}^2. \tag{10}$$

Es fácil ver que, en este caso,

$$P\big((X,Y) \in [a,b] \times \in [c,d]\big) = P\big(X \in [a,b], Y \in [c,d]\big) = P\big(X \in [a,b]\big) \cdot P\big(Y \in [c,d]\big),$$

para todo par de intervalos $[a,b], [c,d] \subset \mathbb{R}$.
El valor esperado de una función h de ambas variables aleatorias ($h : \mathbb{R}^2 \to \mathbb{R}$) viene dado por

$$E\left[h(X,Y)\right] = \iint\limits_{\mathbb{R}^2} h(x,y) f_{X,Y}(x,y)\, dxdy. \tag{11}$$

Sean dos funciones $k, m : \mathbb{R} \to \mathbb{R}$. Luego, si X e Y son variables aleatorias independientes,

$$E\left[k(X)m(Y)\right] = \iint\limits_{\mathbb{R}^2} k(x)m(y) f_X(x) f_Y(y)\, dxdy = E\left[k(X)\right] E\left[m(Y)\right]. \tag{12}$$

Covarianza y correlación

La **covarianza** de dos variables aleatorias está definida por

$$\mathrm{Cov}\left[X,Y\right] = E\left[(X - E\left[X\right])(Y - E\left[Y\right])\right] = E\left[XY\right] - E\left[X\right] E\left[Y\right]. \tag{13}$$

Es fácil ver que, si X e Y son variables aleatorias independientes, entonces $\mathrm{Cov}\left[X,Y\right] = 0$.
Sin embargo, el hecho que $\mathrm{Cov}\left[X,Y\right] = 0$ no implica que X e Y sean independientes.
El **coeficiente de correlación** de dos variables aleatorias está definido como

$$\rho\left(X,Y\right) = \frac{\mathrm{Cov}\left[X,Y\right]}{\sqrt{\mathrm{Var}\left[X\right] \mathrm{Var}\left[Y\right]}}. \tag{14}$$

Se puede demostrar que $\rho(X,Y) = \pm 1$ sii $\exists a, b \in \mathbb{R}$ tales que

$$P\big(Y = aX + b\big) = 1. \tag{15}$$

Es decir, $\rho(X,Y) = \pm 1$ sii existe una relación lineal entre ambas variables aleatorias. Más aún, $\mathrm{sign}(a) = \mathrm{sign}(\rho(X,Y))$.

Suma de dos variables aleatorias

Sean X e Y dos variables aleatorias. Luego,

$$E[X+Y] = E[X] + E[Y]$$
$$E\left[(X+Y)^2\right] = E\left[X^2\right] + 2E[XY] + E\left[Y^2\right],$$

siempre que los momentos en los lados derechos estén definidos. A partir de estas ecuaciones, es fácil ver que

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + 2\mathrm{Cov}[X,Y] + \mathrm{Var}[Y]. \tag{16}$$

En el caso particular de variables independientes, tenemos

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]. \tag{17}$$

Si X e Y son variables aleatorias discretas, entonces

$$P\big(X+Y=z\big) = \sum_{\{(x,y)\in\mathcal{R}_X\times\mathcal{R}_Y:\, x+y=z\}} p_{X,Y}(x,y). \tag{18}$$

En el caso particular en que X e Y independientes,

$$P\big(X+Y=z\big) = \sum_{x\in\mathcal{R}_X} p_X(x)p_Y(z-x) = \sum_{y\in\mathcal{R}_Y} p_Y(y)p_X(z-y). \tag{19}$$

De manera análoga, si X e Y son variables aleatorias continuas e independientes, tenemos

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x)f_Y(z-x)\,dx = \int_{\mathbb{R}} f_Y(y)f_X(z-y)\,dy. \tag{20}$$

Problema 13 de la guía 5

Considere la siguiente tabla de la función de probabilidad conjunta de las variables aleatorias discretas X e Y:

Y\X	**0**	**1**	**2**	**3**	**4**	**5**
0	0	0.01	0.03	0.05	0.07	0.09
1	0.01	0.02	0.04	0.05	0.06	0.08
2	0.01	0.03	0.05	0.05	0.05	0.06
3	0.01	0.02	0.04	0.06	0.06	0.05

Obtener:

1. las distribuciones de probabilidad marginales;
2. la distribución de probabilidades de $X+Y$ y XY;
3. el valor esperado y la varianza de $X+Y$ y de XY;
4. la covarianza $\mathrm{Cov}[X,Y]$.

Respuesta:
Distribuciones marginales: Es sencillo calcular las distribuciones marginales sumando las filas (marginal de Y) y las columnas (marginal de X) de la tabla:

Y\X	**0**	**1**	**2**	**3**	**4**	**5**	p_Y
0	0.00	0.01	0.03	0.05	0.07	0.09	0.25
1	0.01	0.02	0.04	0.05	0.06	0.08	0.26
2	0.01	0.03	0.05	0.05	0.05	0.06	0.25
3	0.01	0.02	0.04	0.06	0.06	0.05	0.24
p_X	0.03	0.08	0.16	0.21	0.24	0.28	—

A partir de esta tabla, es fácil ver que las variables no son independientes. En efecto, tenemos, por ejemplo,

$$p_{X,Y}(0,0) = 0 \neq 0.03 \times 0.25 = p_X(0)p_Y(0). \tag{21}$$

Los valores esperados de cada variable aleatoria se pueden calcular como

$$E[X] = \sum_{x=0}^{5} x p_X(x) = 3.39, \tag{22}$$

$$E[X] = \sum_{y=0}^{3} y p_Y(y) = 1.48, \tag{23}$$

$$E\left[X^2\right] = \sum_{x=0}^{5} x^2 p_X(x) = 13.45, \tag{24}$$

$$E\left[Y^2\right] = \sum_{x=0}^{3} y^2 p_Y(y) = 3.42, \tag{25}$$

$$\text{Var}[X] = (13.45) - (3.39)^2 = 1.9579 \Rightarrow \sigma_X \approx 1.8412, \tag{26}$$

$$\text{Var}[Y] = (3.42) - (1.48)^2 = 1.2296 \Rightarrow \sigma_Y \approx 1.3569. \tag{27}$$

Distribución de XY:

Llamemos $W = XY$. Luego, $\mathcal{R}_Z = \{0, 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15\}$. Para calcular la función de probabilidad de W, es conveniente superponer a la tabla de probabilidades conjuntas la siguiente tabla con los valores de W:

Y\X	**0**	**1**	**2**	**3**	**4**	**5**
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	6	8	10
3	0	3	6	9	12	15

Marcando con distintos colores los distintos casos que nos interesan de la tabla de probabilidades conjuntas:

Y\X	**0**	**1**	**2**	**3**	**4**	**5**
0	0.00	0.01	0.03	0.05	0.07	0.09
1	0.01	0.02	0.04	0.05	0.06	0.08
2	0.01	0.03	0.05	0.05	0.05	0.06
3	0.01	0.02	0.04	0.06	0.06	0.05

Luego,

w	0	1	2	3	4	5	6	8	9	10	12	15
p_W	0.28	0.02	0.07	0.07	0.11	0.08	0.09	0.05	0.06	0.06	0.06	0.05

Los momentos de W:

$$E[W] = \sum_{w \in \mathcal{R}_W} w p_W(w) = 4.76,$$
$$E[W^2] = \sum_{w \in \mathcal{R}_W} w^2 p_W(w) = 41.88,$$
$$\text{Var}[W] = (41.88) - (4.76)^2 = 19.2224 \Rightarrow \sigma_W \approx 4.3843.$$

A partir de los resultados hasta aquí obtenidos, podemos calcular la covarianza de las variables X e Y:

$$\text{Cov}[X, Y] = E[XY] - E[X] E[Y] = -0.2572. \tag{28}$$

El coeficiente de correlación es:

$$\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} \approx -0.1658. \tag{29}$$

Distribución de $X + Y$:

Definamos $Z = X + Y$. Luego, $\mathcal{R}_Z = \{0, 1, 2, \cdots, 8\}$. La función de probabilidad de Z se puede encontrar sumando los valores en las anti-diagonales de la tabla de probabilidades conjuntas, escritas con distintos colores:

Y\X	**0**	**1**	**2**	**3**	**4**	**5**
0	0.00	0.01	0.03	0.05	0.07	0.09
1	0.01	0.02	0.04	0.05	0.06	0.08
2	0.01	0.03	0.05	0.05	0.05	0.06
3	0.01	0.02	0.04	0.06	0.06	0.05

Por tanto,

z	0	1	2	3	4	5	6	7	8
p_Z	0.00	0.02	0.06	0.13	0.19	0.24	0.19	0.12	0.05

Ahora podemos calcular los momentos de Z:

$$E[Z] = \sum_{z=0}^{8} z p_Z(z) = 4.87,$$
$$E[Z^2] = \sum_{z=0}^{8} z^2 p_Z(z) = 26.39,$$
$$\text{Var}[Z] = (26.39) - (4.87)^2 = 2.6731 \Rightarrow \sigma_Z \approx 1.6350.$$

Obsérvese que:

$$E[X] + E[Y] = 3.39 + 1.48 = 4.87 = E[Z],$$
$$\text{Var}[X] + 2\text{Cov}[X, Y] + \text{Var}[Y] = 1.9579 + 2(-0.2572) + 1.2296 = 2.5731 = \text{Var}[Z].$$

Ejercicio 24 de la guía 5

La función de densidad de probabilidad conjunta de (X, Y) es

$$f_{X,Y}(x,y) = \begin{cases} a(x+y) & 0 < x < 3,\, 0 < y < 3 \\ 0 & x \notin (0,3) \times (0,3) \end{cases}$$

1. Calcular $P\big(1 < X < 2, 1 < Y < 2\big)$
2. Calcular $E[X]$, $E[Y]$, σ_X y σ_Y.

Respuesta:
Primero debemos calcular la constante a. Para ello, notemos que

$$\begin{aligned} 1 &= \iint\limits_{\mathbb{R}^2} f_{X,Y}(x,y)\,dxdy = \int_0^3\int_0^3 a(x+y)\,dxdy = \int_0^3 a\left(\frac{x^2}{2}+xy\right)\Big|_{x=0}^{x=3} dy \\ &= \int_0^3 a\left(\frac{9}{2}+3y\right)\,dy = a\left(\frac{9}{2}y+\frac{3}{2}y^2\right)\Big|_{y=0}^{y=3} \\ &= 27a \Rightarrow a = \frac{1}{27}. \end{aligned}$$

La probabilidad solicitada se puede calcular como:

$$\begin{aligned} P\big(1<X<2, 1<Y<2\big) &= \int_1^2\int_1^2 \frac{1}{27}(x+y)\,dxdy = \int_1^2 \frac{1}{27}\left(\frac{x^2}{2}+xy\right)\Big|_{x=1}^{x=2} dy \\ &= \int_1^2 \frac{1}{27}\left(\frac{3}{2}+y\right)\,dy = \frac{1}{27}\left(\frac{3}{2}y+\frac{1}{2}y^2\right)\Big|_{y=1}^{y=2} \\ &= \frac{1}{9}. \end{aligned}$$

También podemos calcular

$$\begin{aligned} E[XY] &= \int_0^3\int_0^3 xy\frac{1}{27}(x+y)\,dxdy = \int_0^3 \frac{1}{27}\left(\frac{x^3}{3}y+\frac{1}{2}x^2y^2\right)\Big|_{x=0}^{x=3} dy \\ &= \int_0^3 \frac{1}{27}\left(9y+\frac{9}{2}y^2\right)\,dy = \frac{1}{27}\left(\frac{9}{2}y^2+\frac{1}{6}y^3\right)\Big|_{y=0}^{y=3} \\ &= \frac{5}{3}. \end{aligned}$$

Densidades marginales:
Las densidad marginal $f_X(x)$ para $x \in (0,3)$ es:

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = \int_0^3 \frac{1}{27}(x+y)\,dy = \frac{1}{27}\left(xy+\frac{1}{2}y^2\right)\Big|_{y=0}^{y=3} \\ &= \frac{1}{6}+\frac{x}{9}. \end{aligned}$$

Por tanto, tenemos que

$$f_X(x) = \begin{cases} \frac{1}{6} + \frac{x}{9} & x \in (0,3) \\ 0 & x \notin (0,3) \end{cases} \tag{30}$$

Es fácil ver que Y tiene la misma densidad que X. Es decir, X e Y están igualmente distribuidas. Por otro lado,

$$f_{X,Y}(1,1) = \frac{1}{27}(1+1) = \frac{2}{27} \neq \left(\frac{1}{6} + \frac{1}{9}\right)\left(\frac{1}{6} + \frac{1}{9}\right) = f_X(1)f_Y(1), \tag{31}$$

por lo que X e Y no son independientes.

Los valores esperados son:

$$\begin{aligned} E\left[Y\right] = E\left[X\right] &= \int_0^3 x\left(\frac{1}{6} + \frac{x}{9}\right)\, dx = \left.\left(\frac{x^2}{12} + \frac{x^3}{27}\right)\right|_{x=0}^{x=3} \\ &= \frac{7}{4}, \\ E\left[Y^2\right] = E\left[X^2\right] &= \int_0^3 x^2\left(\frac{1}{6} + \frac{x}{9}\right)\, dx = \left.\left(\frac{x^3}{18} + \frac{x^4}{36}\right)\right|_{x=0}^{x=3} \\ &= \frac{15}{4}, \\ \mathrm{Var}\left[Y\right] = \mathrm{Var}\left[X\right] &= \frac{15}{4} - \left(\frac{7}{4}\right)^2 = \frac{11}{16} \Rightarrow \sigma_Y = \sigma_X \approx 0.8292. \end{aligned}$$

Con estos resultados, podemos calcular la covarianza de X e Y:

$$\mathrm{Cov}\left[X,Y\right] = E\left[XY\right] - E\left[X\right]E\left[Y\right] = \frac{5}{3} - \frac{7}{4}\cdot\frac{7}{4} = -\frac{67}{48}. \tag{32}$$

Dado que la covarianza es distinta de cero, podemos afirmar que las variables aleatorias no son independientes.