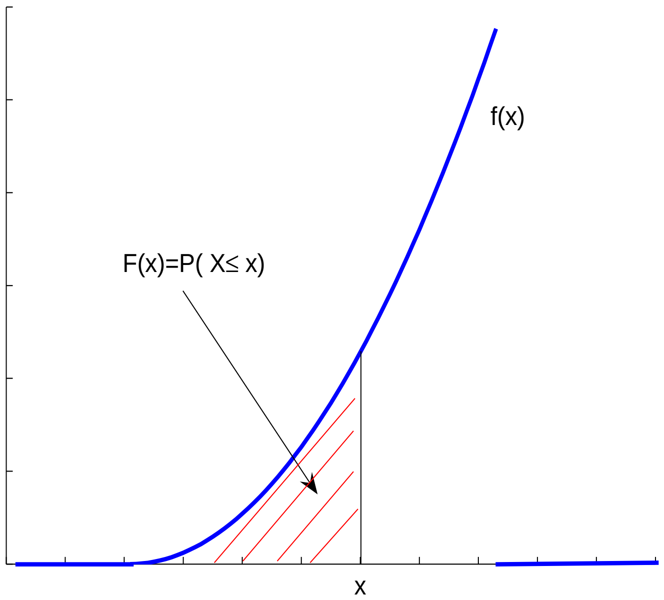
Introducción a las distribuciones multivariantes

En determinadas situaciones el resultado de un experimento aleatorio no es un solo valor numérico X, sinó una colección de resultados $X_1, \ldots, X_n$. Por ejemplo, sobre un individuo podemos medir su altura X y su peso Y, en un árbol podemos medir la altura X y el perímetro de su tronco Y, el resultado de un análisis de sangre es un conjunto de medidas de concentraciones de ciertas substancias $X_1, \ldots, X_n$.

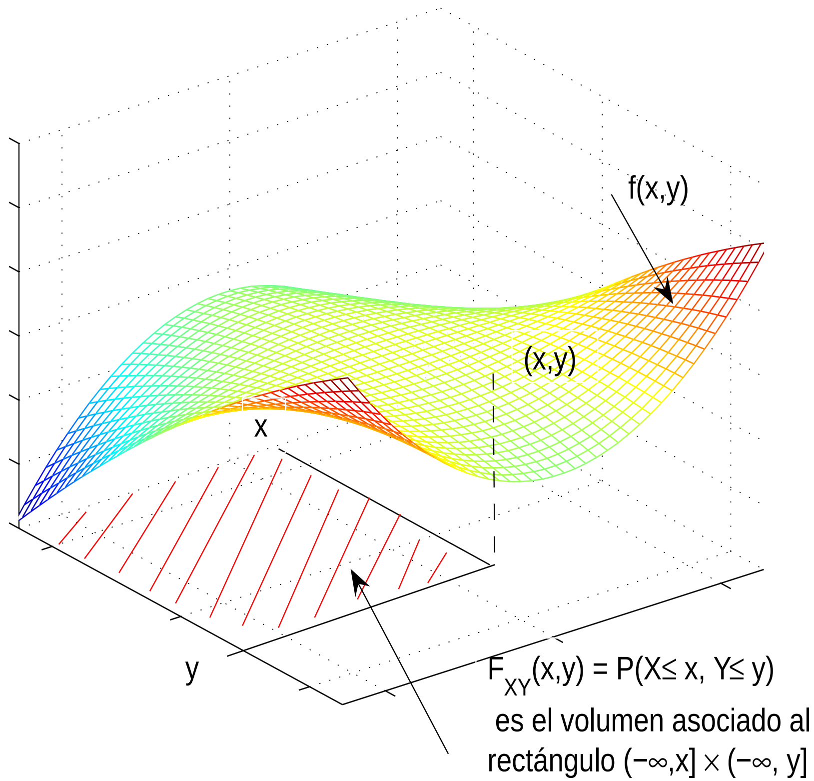
Diremos que $\mathbf{X} = (X_1, \ldots, X_n)$ es un vector aleatorio si cada una de sus componentes es una variable aleatoria. Nos restringiremos al caso $n = 2$ y escribiremos (X, Y).

La función de **distribución conjunta** del vector aleatorio (X, Y) es

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = P((X, Y) \in (-\infty, x] \times (-\infty, y])$$



caso univariante



caso bivariante

La función de distribución de un vector aleatorio tiene las siguientes propiedades:

1. $0 \leq F_{XY}(x, y) \leq 1$,
2. $F_{XY}(x, y)$ es creciente respecto de x y respecto de y,
3. $\lim_{x,y \to -\infty} F_{XY}(x, y) = 0$,
   $\lim_{x \to -\infty} F_{XY}(x, y) = 0$, $\lim_{y \to -\infty} F_{XY}(x, y) = 0$,
4. $\lim_{x \to +\infty} F_{XY}(x, y) = F_Y(y)$ es la **distribución marginal** de Y,
   $\lim_{y \to +\infty} F_{XY}(x, y) = F_X(x)$ es la **distribución marginal** de X
5. $\lim_{x,y \to +\infty} F_{XY}(x, y) = 1$.

9.1. Vectores aleatorios discretos

El vector aleatorio (X, Y) se denomina **discreto** si solamente puede tomar un conjunto finito o infinito numerable de valores $\{(x_i, y_j)\}$, es decir, si les variables aleatorias X, Y son variables aleatorias discretas.

Si $p_{ij} = P(X = x_i, Y = y_j)$, entonces se define la **función de probabilidad conjunta** $f(x, y)$ como

$$f(x,y) = \begin{cases} p_{ij} = P(X = x_i, Y = y_j), & \text{si } (x,y) = (x_i, y_j), \\ 0, & \text{en otro caso.} \end{cases}$$

Se definen las **probabilidades marginales** de X como

$$p_{i\bullet} = P(X = x_i) = \sum_j p_{ij}$$

y las **probabilidades marginales** de Y como

$$p_{\bullet j} = P(Y = y_j) = \sum_i p_{ij}.$$

Se cumple la siguiente propiedad:

$$\sum_i p_{i\bullet} = \sum_j p_{\bullet j} = \sum_{i,j} p_{ij} = 1.$$

Ejemplo: *Sean X e Y dos v.a. discretas que sólo pueden tomar los valores -1, 0 y 1. La siguiente tabla contiene las probabilidades conjuntas del vector (X, Y):*

		X		
		-1	0	1
Y	-1	2/16	2/16	1/16
	0	3/16	0	3/16
	1	1/16	3/16	1/16

Las probabilidades marginales de X y de Y son

X	$p_{i\bullet}$
-1	6/16
0	5/16
1	5/16
	1

Y	$p_{\bullet j}$
-1	5/16
0	6/16
1	5/16
	1

9.2. Vectores aleatorios continuos

El vector aleatorio (X, Y) se denomina **continuo** si existe una **función de densidad conjunta** f_{XY} tal que la función de distribución conjunta puede expresarse como

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)\,du\,dv, \quad (x,y) \in \mathbb{R}^2.$$

La función de densidad conjunta cumple las siguientes propiedades:

1. $f_{XY} \geq 0$,
2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx\,dy = 1$,
3. $f_{XY}(u,v) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}\Big|_{(x,y)=(u,v)}$,
4. si $D \subset \mathbb{R}^2$, $P((X,Y) \in D) = \int\int_D f_{XY}(x,y)\,dx\,dy$.

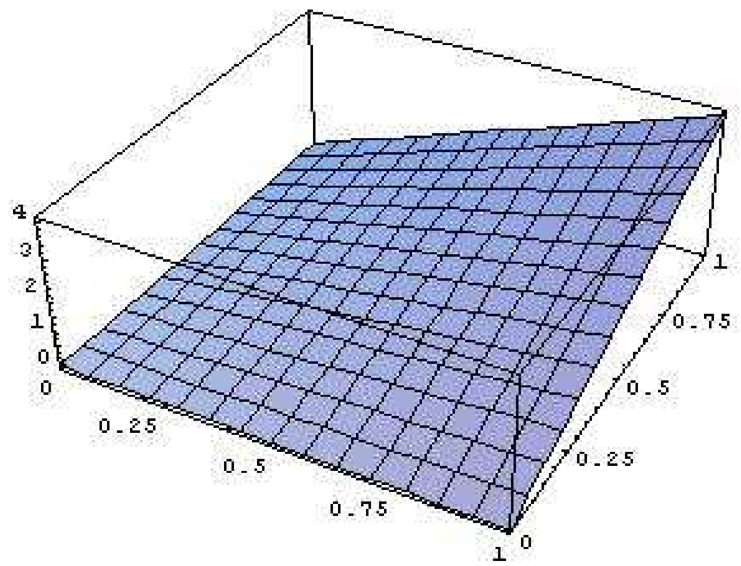
Las **funciones de densidad marginales** son

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx.$$

Ejemplo 55 *Sea* (X, Y) *un vector aleatorio continuo con función de densidad conjunta*

$$f_{XY}(x, y) = \begin{cases} 4xy, & 0 \leq x, y \leq 1, \\ 0, & \textit{en otro caso.} \end{cases}$$

Hallad las densidades marginales de X *y de* Y*. Calculad las probabilidades* $P(X \leq 1/2, Y \leq 1/3)$ *y* $P(1/2 \leq X \leq 1)$*.*



Solución: *La función de densidad de X se obtiene intergando la densidad conjunta respecto de y, es decir:*

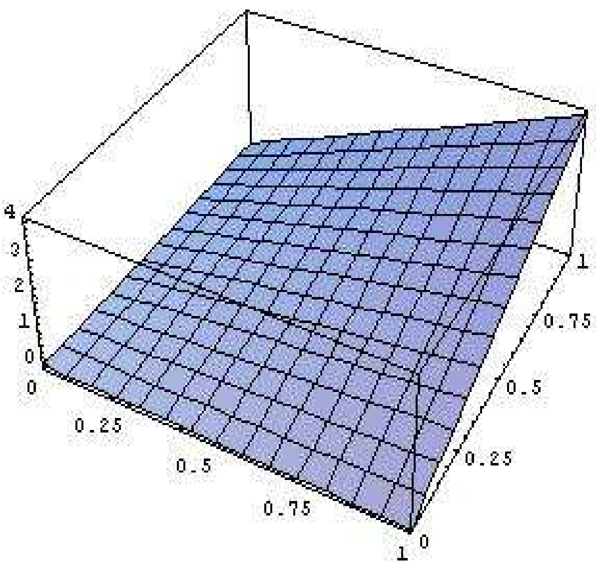
$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy = \int_0^1 4xy\,dy = 4x\frac{y^2}{2}\Big|_0^1 = 2x.$$

La función de densidad de X es

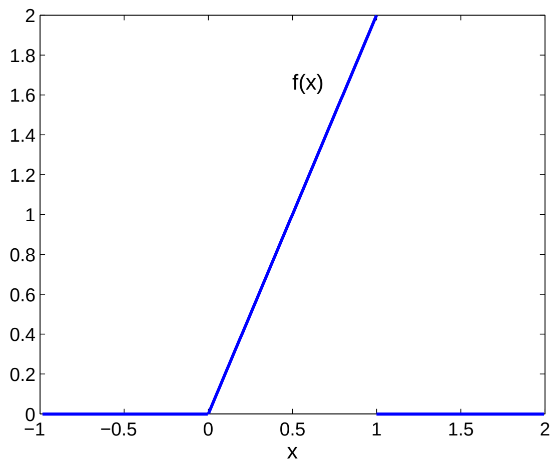
$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1, \\ 0, & \textit{en otro caso.} \end{cases}$$

y, por simetría, la función de densidad de Y es

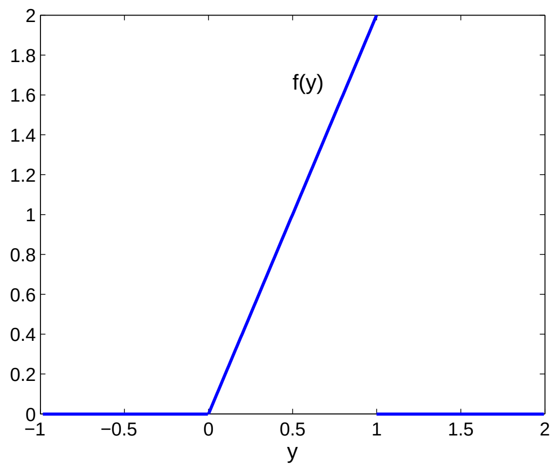
$$f_Y(y) = \begin{cases} 2y, & 0 \le y \le 1, \\ 0, & \textit{en otro caso.} \end{cases}$$



densidad conjunta del vector (X, Y)



densidad marginal de X



densidad marginal de Y

$$\begin{aligned} P(X \leq 1/2, Y \leq 1/3) &= \int_0^{1/2} \left(\int_0^{1/3} 4xy\, dy \right) dx = \int_0^{1/2} \left(4x\, \frac{y^2}{2} \right) \Big|_0^{1/3} \\ &= \int_0^{1/2} 2x\, \frac{1}{9}\, dx = \left(\frac{2}{9}\, \frac{x^2}{2} \right) \Big|_0^{1/2} = \frac{1}{9}\, \frac{1}{4} = \frac{1}{36}. \end{aligned}$$

$$P(1/2 \leq X \leq 1) = \int_{1/2}^{1} f_X(x)\, dx = \int_{1/2}^{1} 2x\, dx = \left(2\, \frac{x^2}{2} \right) \Big|_{1/2}^{1} = 1 - \frac{1}{4} = \frac{3}{4}.$$

Ejemplo 56 *Sea* (X, Y) *un vector aleatorio continuo con función de densidad conjunta*

$$f_{XY}(x, y) = \begin{cases} 2, & 0 \leq x \leq 1, 0 \leq y \leq x, \\ 0, & \textit{en otro caso.} \end{cases}$$

Hallad las densidades marginales de X *y de* Y*. Calculad las probabilidades* $P(X \geq 1/2, Y \geq 1/2)$ *y* $P(Y \geq 1/3)$*.*

Solución:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dy = \int_0^x 2\, dy = 2x, \quad 0 \leq x \leq 1.$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dx = \int_y^1 2\, dy = 2(1 - y), \quad 0 \leq y \leq 1.$$

$$P(X \geq 1/2, Y \geq 1/2) = \int_{1/2}^{1} \left(\int_{1/2}^{x} 2\, dy \right) dx = \int_{1/2}^{1} (2y)|_{1/2}^{x}\, dx$$

$$= \int_{1/2}^{1} (2x - 1)\, dx = \left(\frac{2x^2}{2} - x \right) \Big|_{1/2}^{1} = \frac{1}{4}.$$

$$P(Y \geq 1/3) = \int_{1/3}^{1} f_Y(y)\, dy = \int_{1/3}^{1} 2(1-y)\, dy = 2 \left(y - \frac{y^2}{2} \right) \Big|_{1/3}^{1} = 0.4444.$$

Ejemplo 57 *Se lanzan dos dados y se denota por X el número de unos obtenidos y por Y el número de puntuaciones impares. Hallad las funciones de probabilidad conjunta y marginales.*

Solución: *X e Y sólo pueden tomar valores 0,1,2. Las probabilidades conjuntas del vector (X, Y) son:*

$p_{11} = P(X = 0, Y = 0) = 3/6 \cdot 3/6 = 1/4,$

$p_{21} = P(X = 1, Y = 0) = 0,$

$p_{31} = P(X = 2, Y = 0) = 0,$

$p_{12} = P(X = 0, Y = 1) = 2/6 \cdot 3/6 + 3/6 \cdot 2/6 = 1/3,$

$p_{13} = P(X = 0, Y = 2) = 2/6 \cdot 2/6 = 1/9,$

$p_{22} = P(X = 1, Y = 1) = 1/6 \cdot 3/6 + 3/6 \cdot 1/6 = 1/6,$

$p_{23} = P(X = 1, Y = 2) = 1/6 \cdot 2/6 + 2/6 \cdot 1/6 = 1/9,$

$p_{32} = P(X = 2, Y = 1) = 0,$

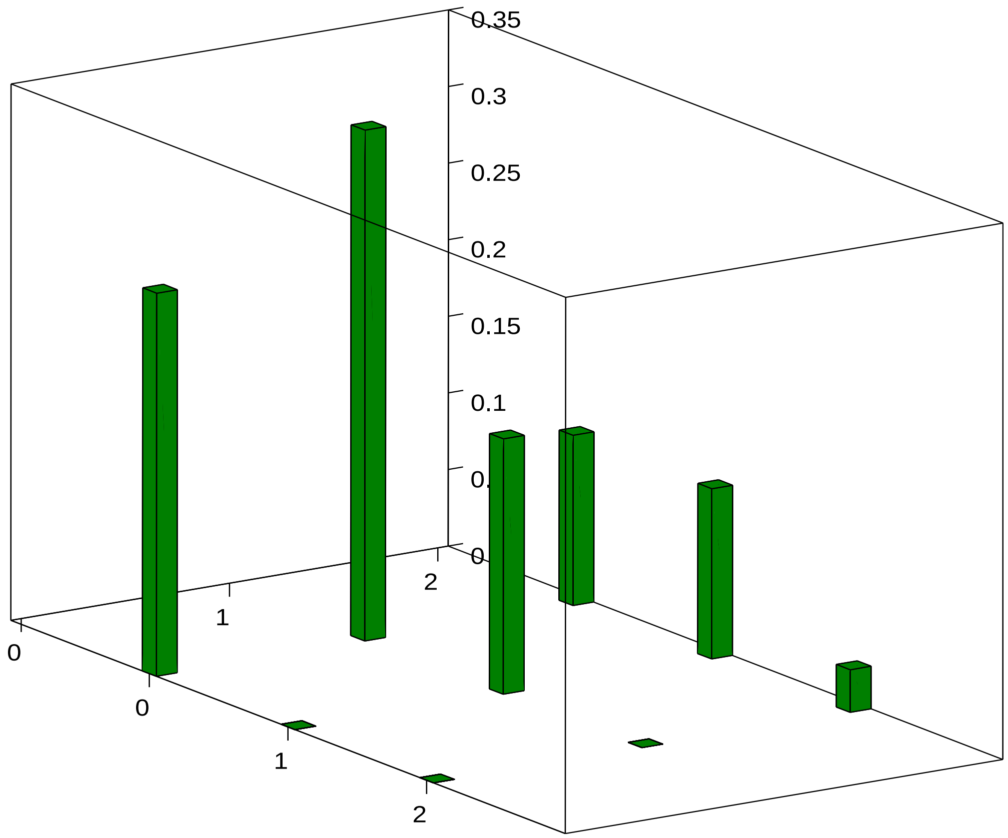
$p_{33} = P(X = 2, Y = 2) = 1/6 \cdot 1/6 = 1/36.$

Podemos representar las probabilidades conjuntas y marginales en la siguiente tabla:

		Y 0	1	2	$p_{i\bullet}$
X	0	1/4	1/3	1/9	25/35
	1	0	1/6	1/9	5/18
	2	0	0	1/36	1/36
	$p_{\bullet j}$	1/4	1/2	1/4	1

donde $X =$ *"número de unos",* $Y =$ *"número de puntuaciones impares" al lanzar dos dados.*

Figura 1: Función de probabilidad conjunta de las v.a. del Ejemplo 57



9.3. Independencia. Covarianza y correlación.

Las v.a. X e Y se dice que son **independientes** si

$$F_{XY}(x,y) = F_X(x)\,F_Y(y), \quad \forall x, y. \tag{1}$$

Si X, Y son v.a. continuas, la definición (1) es equivalente a

$$f_{XY}(x,y) = f_X(x)\,f_Y(y), \quad \forall x, y.$$

Si X, Y son v.a. discretas, la definición (1) es equivalente a

$$p_{ij} = p_{i\bullet}\,p_{\bullet j}, \quad \forall i, j.$$

Ejemplos:

En el ejemplo 55 las v.a. X *e* Y *son independientes, puesto que*

$$f_{XY}(x,y) = 4xy = 2x\,2y = f_X(x)\,f_Y(y), \quad 0 \leq x, y \leq 1.$$

En el ejemplo 57 las v.a. X *e* Y *no son independientes, puesto que, por ejemplo:*

$$p_{11} = P(X=0, Y=0) = \frac{1}{4} \neq p_{1\bullet}\,p_{\bullet 1} = \frac{25}{36} \times \frac{1}{4} = \frac{25}{144}.$$

Ejercicio: *Comprovad que las v.a.* X, Y *del ejemplo 56 no son independientes.*

Para medir la relación entre las variables X e Y se introduce la **covarianza** entre X e Y como

$$\mathsf{cov}(X,Y) = \mathsf{E}((X - m_X)(Y - m_Y)) = \mathsf{E}(X\,Y) - \mathsf{E}(X)\ \mathsf{E}(Y),$$

donde $m_X = \mathsf{E}(X)$, $m_Y = \mathsf{E}(Y)$,

$$\mathsf{E}(X\,Y) = \begin{cases} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,y\,f_{XY}(x,y)\,dx\,dy, & \text{si } X, Y \text{ son continuas,} \\ \sum_i \sum_j x_i\,y_j\,p_{ij}, & \text{si } X, Y \text{ son discretas.} \end{cases}$$

El **coeficiente de correlación** (lineal) entre X e Y es una *normalización de la covarianza* y se define como

$$\rho_{XY} = \frac{\mathsf{cov}(X,Y)}{\sqrt{\mathsf{var}(X)\ \mathsf{var}(Y)}}$$

y tiene las siguientes propiedades:

1. $-1 \leq \rho_{XY} \leq 1$.
2. Si X e Y son independientes entonces $\mathsf{cov}(X,Y) = 0$ y, por tanto, $\rho_{XY} = 0$. Esto se debe a que si X e Y son independientes $\mathsf{E}(X\,Y) = \mathsf{E}(X)\ \mathsf{E}(Y)$.
3. $\mathsf{var}(X+Y) = \mathsf{var}(X) + \mathsf{var}(Y) + 2\ \mathsf{cov}(X,Y)$.
4. $\mathsf{var}(X-Y) = \mathsf{var}(X) + \mathsf{var}(Y) - 2\ \mathsf{cov}(X,Y)$.

Ejemplo 58 *Calcular la covarianza entre las variables X e Y de los ejemplos 55, 56 y 57.*

Solución: *En el ejemplo 55, $\mathsf{cov}(X,Y)=0$, puesto que X, Y son independientes (y por tanto, $\mathsf{E}(X\,Y)=\mathsf{E}(X)\ \mathsf{E}(Y)$).*

En el ejemplo 56, tenemos que

$$\mathsf{E}(X)=\int_{-\infty}^{+\infty} x\, f_X(x)\,dx=\int_0^1 x\,2x\,dx=2/3,$$

$$\mathsf{E}(Y)=\int_{-\infty}^{+\infty} y\, f_Y(y)\,dy=\int_0^1 y\,2(1-y)\,dy=1/3,$$

$$\mathsf{E}(X\,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f_{XY}(x,y)\,dx\,dy=\int_0^1\left(\int_0^x 2xy\,dy\right)dx=1/4,$$

$$\mathsf{cov}(X,Y)=\mathsf{E}(X\,Y)-\mathsf{E}(X)\ \mathsf{E}(Y)=1/4-2/3\cdot 1/3=1/36.$$

En el ejemplo 57, tenemos que

		Y 0	1	2	$p_{i\bullet}$
X	0	1/4	1/3	1/9	25/35
	1	0	1/6	1/9	5/18
	2	0	0	1/36	1/36
	$p_{\bullet j}$	1/4	1/2	1/4	1

$$\begin{aligned}
\mathsf{E}(X) &= \sum_i x_i\, p_{i\bullet} = 0 \cdot 25/36 + 1 \cdot 5/18 + 2 \cdot 1/36 = 1/3, \\
\mathsf{E}(Y) &= \sum_j y_j\, p_{\bullet j} = 0 \cdot 1/4 + 1 \cdot 1/2 + 2 \cdot 1/4 = 1, \\
\mathsf{E}(X\,Y) &= \sum_i \sum_j x_i\, y_j\, p_{ij} = 1 \cdot 1 \cdot 1/6 + 1 \cdot 2 \cdot 1/9 + 2 \cdot 2 \cdot 1/36 = 1/2, \\
\mathsf{cov}(X,Y) &= \mathsf{E}(X\,Y) - \mathsf{E}(X)\ \mathsf{E}(Y) = 1/2 - 1/3 \cdot 1 = 1/6.
\end{aligned}$$

Ejercicio: *Calculad* ρ_{XY}.

Ejercicio:

Considerad la distribución bivariante resumida en la tabla siguiente:

	y_j			
x_i	10	20	30	$p_{i\bullet}$
25	0.08	0.56	0.12	
45	0.12	0	0.12	
$p_{\bullet j}$				

(a) Calculad las esperanzas $\mathsf{E}(X)$, $\mathsf{E}(Y)$ y la covarianza $\mathsf{cov}(X,Y)$.
(b) ¿Son X y Y independientes?
(c) Calculad el coeficiente de correlación ρ_{XY}.

9.4. La distribución normal bivariante

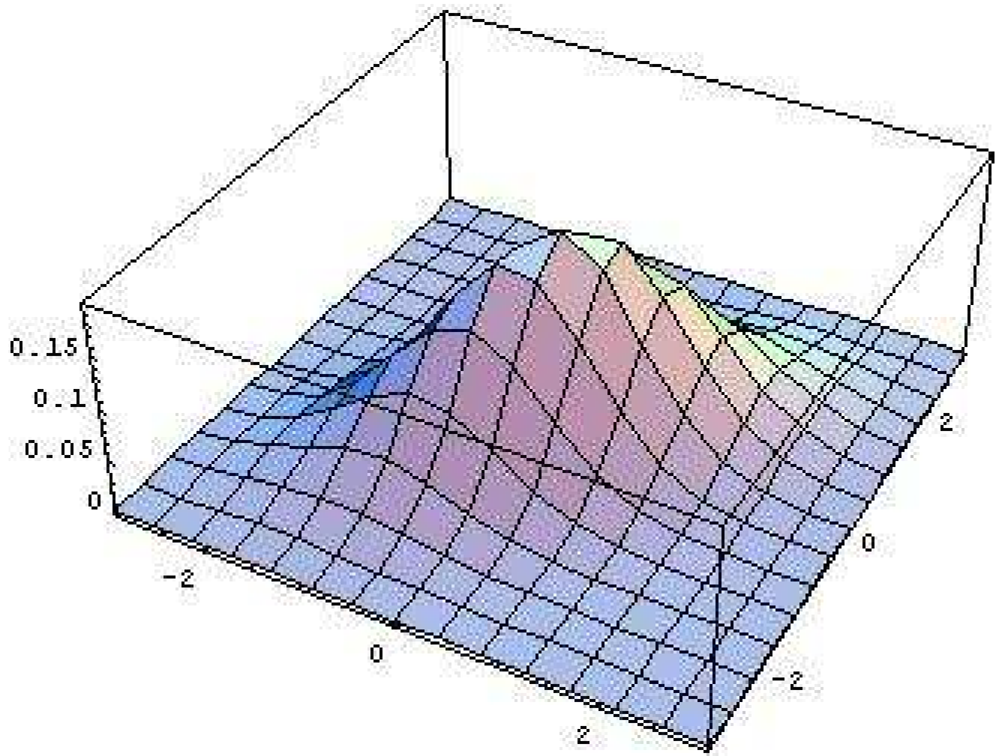
Se dice que un vector aleatorio (X, Y) tiene una distribución normal bidimensional (o bivariante) si es un vector aleatorio continuo con función de densidad conjunta

$$f_{XY}(x, y) = \frac{1}{2\pi\, \sigma_X\, \sigma_Y\, \sqrt{1-\rho^2}}$$

$$\cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2} - 2\rho\frac{(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right]\right\},$$

donde $m_X = \mathsf{E}(X)$, $m_Y = \mathsf{E}(Y)$, $\sigma_X^2 = \mathsf{var}(X)$, $\sigma_Y^2 = \mathsf{var}(Y)$ y $\rho = \mathsf{cov}(X, Y)/(\sigma_X\sigma_Y)$ es el coeficiente de correlación entre X e Y.

Función de densidad de una ley normal bivariante



Observad que si $\rho = 0$, entonces:

$$
\begin{aligned}
f_{XY}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2}\left[\frac{(x-m_X)^2}{\sigma_X^2}+\frac{(y-m_Y)^2}{\sigma_Y^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{1}{2}\frac{(x-m_X)^2}{\sigma_X^2}\right\}\cdot\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{1}{2}\frac{(y-m_Y)^2}{\sigma_Y^2}\right\} \\
&= f_X(x)\cdot f_Y(y),
\end{aligned}
$$

donde $f_X(x)$ y $f_Y(y)$ son las funciones de densidad de dos variables aleatorias normales $N(m_X,\sigma_X^2)$ y $N(m_Y,\sigma_Y^2)$, respectivamente.

Acabamos de demostrar una propiedad muy importante de las leyes normales:
Si un vector aleatorio (X,Y) es normal bivariante y el coeficiente de correlación $\rho = 0$, entonces las v.a. marginales X e Y son también normales e independientes.

Es decir, si (X,Y) es normal bivariante el concepto de *incorrelación* entre v.a. equivale al concepto de *independencia* de v.a.