

Probabilidad y Estadística

Grado en Ingeniería Informática

Tema 4 Vectores aleatorios

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid

javier.carcamo@uam.es

Descripción del tema

1. Vectores aleatorias.
2. Función de distribución conjunta.
3. Vectores discretos.
4. Vectores continuos.
5. Ejemplos de distribuciones discretas multidimensionales.
6. Ejemplos de distribuciones continuas multidimensionales.
7. Distribuciones condicionadas.
8. Variables aleatorias independientes.
9. Distribución de la suma de variables.

Objetivos principales

- Comprender la noción de vector aleatorio.
- Saber calcular distribuciones marginales y condicionadas.

1. Vectores aleatorios

En el tema anterior, dado un experimento aleatorio ϵ , prestamos atención a una sola característica de ϵ considerando una variable $X : \Omega \rightarrow \mathbb{R}$.

Sin embargo, nos puede interesar estudiar simultáneamente varias características del experimento. Es decir, estudiar varias variables a la vez: $X : \Omega \rightarrow \mathbb{R}$, $Y : \Omega \rightarrow \mathbb{R}$, $Z : \Omega \rightarrow \mathbb{R}$, ...

Estudiar las variables simultáneamente proporciona más información que su estudio por separado ya que permite analizar las interrelaciones entre las características involucradas.

Por ejemplo, al seleccionar una persona se puede observar su peso y altura. Por tanto, podemos analizar dos variables por separado o estudiar el comportamiento de las dos variables a la vez. Al estudiar estas variables de manera conjunta podemos observar la relación existente entre ambas. (En general, es razonable esperar que a mayor altura de una persona, su peso será mayor. Esto no lo podríamos apreciar estudiando las variables "peso" y "altura" de forma independiente.)

Otra situación en la que es importante el estudio de muchas características simultáneamente es la realización de una encuesta (por ejemplo electoral).

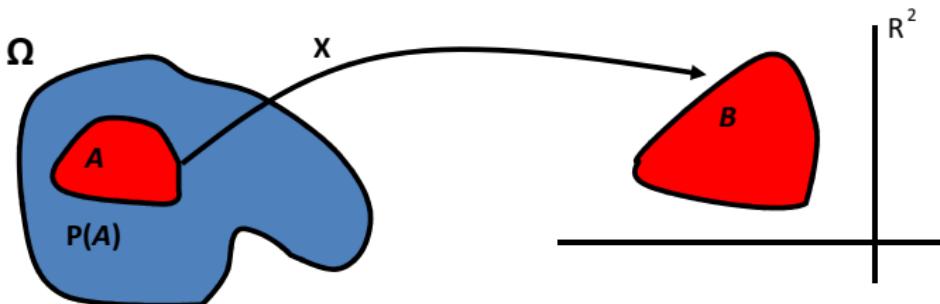
Dado (Ω, \mathcal{F}, P) un espacio de probabilidad, se llama **vector aleatorio** (d -dimensional) a toda aplicación:

$$\begin{aligned}\mathbf{X} = (X_1, \dots, X_d) : \Omega &\longrightarrow \mathbb{R}^d \quad (\text{o } \overline{\mathbb{R}}^d) \\ \omega &\longmapsto \mathbf{X}(\omega) = (X_1(\omega), \dots, X_d(\omega)).\end{aligned}$$

Observación: Realmente, para que \mathbf{X} sea un v.a. debe cumplir una condición de *medibilidad*: Para todo $a_1, \dots, a_d \in \mathbb{R}^d$, se pide que el conjunto $\{\omega \in \Omega : X_1(\omega) \leq a_1, \dots, X_d(\omega) \leq a_d\} \in \mathcal{F}$. Cuando $\mathcal{F} = \mathcal{P}(\Omega)$, esto siempre se cumple.

Idea: Como P es medida de probabilidad sobre (Ω, \mathcal{F}) , sabemos “calcular probabilidades” en Ω vía P . Al tener una aplicación $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^d$ vamos tener una forma de “medir” en \mathbb{R}^d .

La medida de probabilidad P sobre (Ω, \mathcal{F}) induce una *nueva* medida de probabilidad sobre \mathbb{R}^d , $P_{\mathbf{X}}$, que llamaremos **distribución de probabilidad** del vector aleatorio \mathbf{X} .



Notación: $\{\mathbf{X} \in B\} = \mathbf{X}^{-1}(B) = \{\omega \in \Omega : \mathbf{X}(\omega) \in B\}, \quad B \subset \mathbb{R}^d.$

$\mathbf{X} = (X_1, \dots, X_d) : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}^d$ v.a. Consideramos

$$P_{\mathbf{X}} : \mathcal{B}(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d) \longrightarrow [0, 1]$$

$$B \longmapsto P_{\mathbf{X}}(B) = P(\mathbf{X} \in B).$$

$P_{\mathbf{X}}$ se llama **distribución de probabilidad** de \mathbf{X} o **distribución conjunta de las variables** X_1, \dots, X_d y es una medida de probabilidad en $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Nota: $\mathcal{B}(\mathbb{R}^d)$ es una σ -álgebra sobre \mathbb{R}^d (σ -álgebra de Borel) que incluye todos los conjuntos que nos pueden interesar para calcular probabilidades.

Si $\mathbf{X} = (X_1, \dots, X_d) : \Omega \longrightarrow \mathbb{R}^d$ es un vector aleatorio, las proyecciones,

$$\begin{aligned} X_i : \Omega &\longrightarrow \mathbb{R} \quad (\text{o } \overline{\mathbb{R}}) \\ \omega &\longmapsto X_i(\omega), \end{aligned}$$

son variables aleatorias. Las variables X_1, \dots, X_d se llaman **variables marginales** del vector aleatorio \mathbf{X} . Las d distribuciones de probabilidad (sobre \mathbb{R}) asociadas, P_{X_1}, \dots, P_{X_d} , se denominan **distribuciones marginales** de \mathbf{X} .

Nota importante: En general, si conocemos $P_{\mathbf{X}}$ podemos calcular las distribuciones marginales P_{X_1}, \dots, P_{X_d} . Sin embargo, el recíproco *no* es cierto. Es decir, (salvo en algunos casos especiales) aunque sepamos la distribución de X_1, \dots, X_d no podemos calcular la distribución conjunta del vector $\mathbf{X} = (X_1, \dots, X_d)$.

Ejemplo: ϵ : lanzar una moneda al aire 3 veces.

$$\Omega = \{(C, C, C), (C, C, +), \dots, (+, +, +)\}.$$

(Ω, \mathcal{F}, P) : modelo de Laplace (equiprobabilidad). $\text{Card}(\Omega) = 8$.

$$P(\{(C, C, C)\}) = P(\{(C, C, +)\}) = \dots = P(\{(+, +, +)\}) = 1/8.$$

Consideramos $\mathbf{X} = (X, Y) : \Omega \longrightarrow \mathbb{R}^2$, de forma que:

- $X \equiv$ número de caras obtenidas.
- $Y \equiv$ máximo número de caras seguidas obtenidas.

Problema: Calcular la distribución de probabilidad conjunta de (X, Y) y las distribuciones marginales de las variables X e Y .

Primero tenemos que ver qué valores puede tomar el vector \mathbf{X} .

Ejemplo: ϵ : lanzar una moneda al aire 3 veces.

$$\Omega = \{(C, C, C), (C, C, +), \dots, (+, +, +)\}.$$

- $X \equiv$ número de caras obtenidas al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas obtenidas.

Valores que toma el vector (X, Y)

$(X, Y) : \Omega \longrightarrow \mathbb{R} \times \mathbb{R}$	$(X, Y) : \Omega \longrightarrow \mathbb{R} \times \mathbb{R}$
$(C, C, C) \longmapsto (3, 3)$	$(+, +, +) \longmapsto (0, 0)$
$(C, C, +) \longmapsto (2, 2)$	$(C, +, +) \longmapsto (1, 1)$
$(+, C, C) \longmapsto (2, 2)$	$(+, C, +) \longmapsto (1, 1)$
$(C, +, C) \longmapsto (2, 1)$	$(+, +, C) \longmapsto (1, 1)$

(X, Y) toma los valores $\{(0, 0), (1, 1), (2, 1), (2, 2), (3, 3)\}$. Para calcular la distribución de probabilidad de (X, Y) tenemos que calcular la probabilidad de que (X, Y) tome estos 5 valores.

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

$$\begin{aligned} P_{(X,Y)}(\{(0,0)\}) &= P(X = 0, Y = 0) = P(\{\omega \in \Omega : X(\omega) = 0, Y(\omega) = 0\}) \\ &= P(\{(+, +, +)\}) = 1/8. \end{aligned}$$

$$\begin{aligned} P_{(X,Y)}(\{(1,1)\}) &= P(X = 1, Y = 1) = P(\{\omega \in \Omega : X(\omega) = 1, Y(\omega) = 1\}) \\ &= P(\{(C, +, +), (+, C, +), (+, +, C)\}) = 3/8. \end{aligned}$$

$$\begin{aligned} P_{(X,Y)}(\{(2,1)\}) &= P(X = 2, Y = 1) = P(\{\omega \in \Omega : X(\omega) = 2, Y(\omega) = 1\}) \\ &= P(\{(C, +, C)\}) = 1/8. \end{aligned}$$

$$\begin{aligned} P_{(X,Y)}(\{(2,2)\}) &= P(X = 2, Y = 2) = P(\{\omega \in \Omega : X(\omega) = 2, Y(\omega) = 2\}) \\ &= P(\{(C, C, +), (+, C, C)\}) = 2/8. \end{aligned}$$

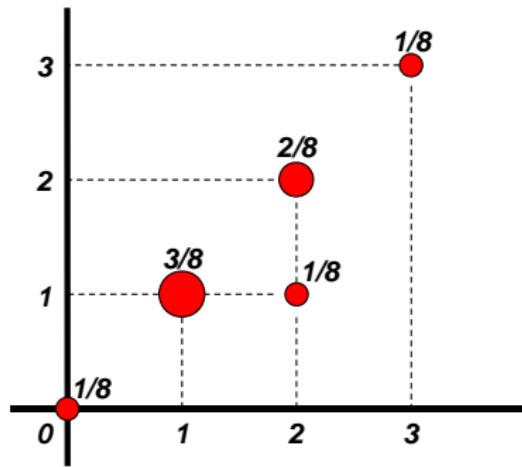
$$\begin{aligned} P_{(X,Y)}(\{(3,3)\}) &= P(X = 3, Y = 3) = P(\{\omega \in \Omega : X(\omega) = 3, Y(\omega) = 3\}) \\ &= P(\{(C, C, C)\}) = 1/8. \end{aligned}$$

El vector (X, Y) toma los valores $(0,0), (1,1), (2,1), (2,2), (3,3)$ con probabilidades respectivas $1/8, 3/8, 1/8, 2/8$ y $1/8$.

1. Vectores aleatorios

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

El vector (X, Y) toma los valores $(0, 0), (1, 1), (2, 1), (2, 2), (3, 3)$ con probabilidades respectivas $1/8, 3/8, 1/8, 2/8$ y $1/8$.

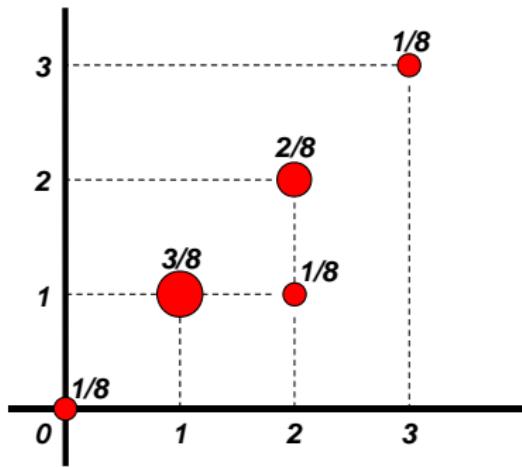


$X \setminus Y$	0	1	2	3
0	1/8	0	0	0
1	0	3/8	1/8	0
2	0	0	2/8	0
3	0	0	0	1/8

1. Vectores aleatorios

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

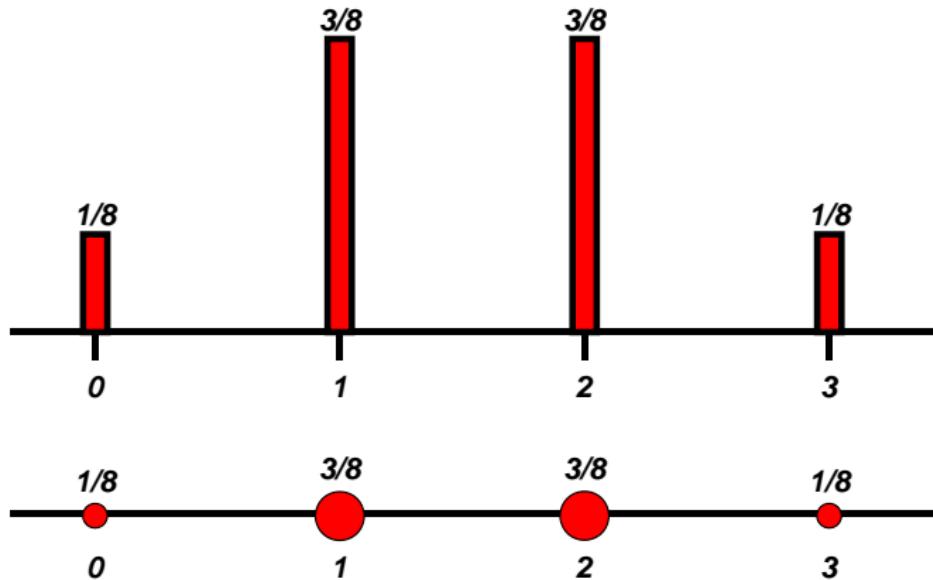
El vector (X, Y) toma los valores $(0, 0), (1, 1), (2, 1), (2, 2), (3, 3)$ con probabilidades respectivas $1/8, 3/8, 1/8, 2/8$ y $1/8$.



$X \setminus Y$	0	1	2	3
0	1/8	0	0	0
1	0	3/8	1/8	0
2	0	0	2/8	0
3	0	0	0	1/8

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

Distribución marginal de X



- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

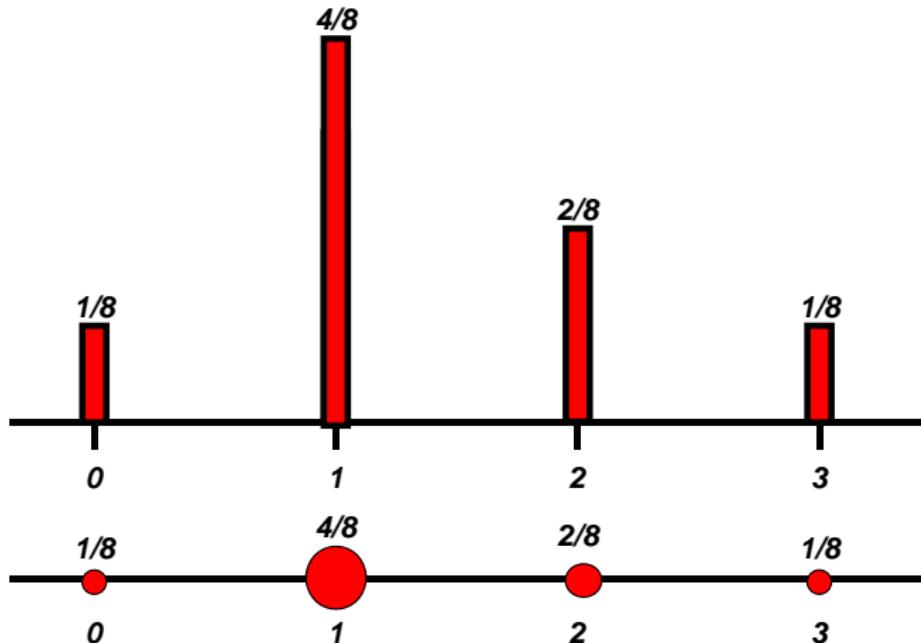
Distribución marginal de Y

$Y : \Omega \rightarrow \mathbb{R}$	$Y : \Omega \rightarrow \mathbb{R}$
$(C, C, C) \mapsto 3$	$(+, +, +) \mapsto 0$
$(C, C, +) \mapsto 2$	$(C, +, +) \mapsto 1$
$(+, C, C) \mapsto 2$	$(+, C, +) \mapsto 1$
$(C, +, C) \mapsto 1$	$(+, +, C) \mapsto 1$

La variable aleatoria Y toma los valores 0,1,2 y 3 con probabilidades respectivas $1/8$, $4/8$, $2/8$ y $1/8$.

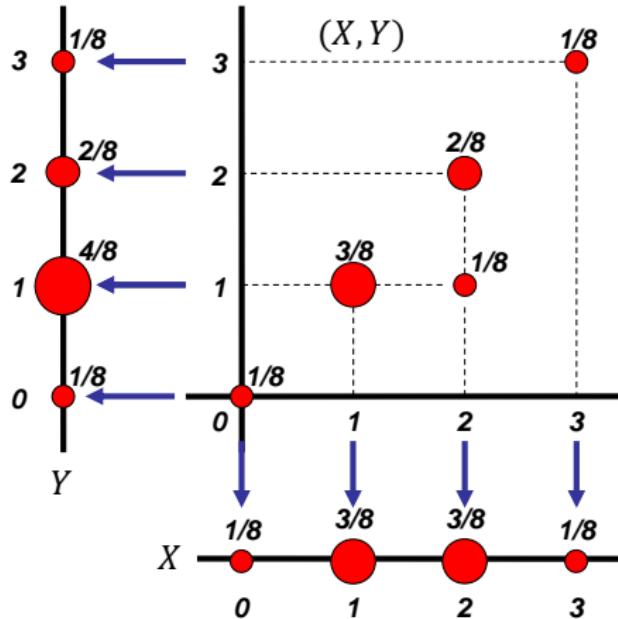
- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

Distribución marginal de Y



- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

Distribución conjunta y marginales de (X, Y)

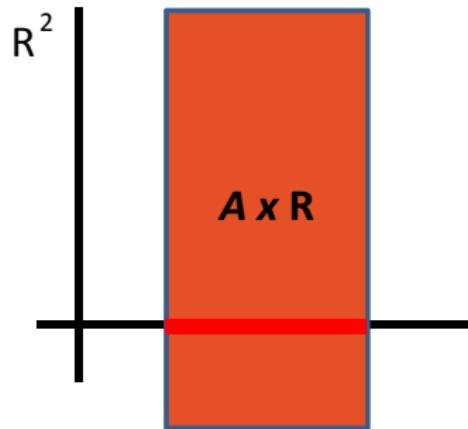
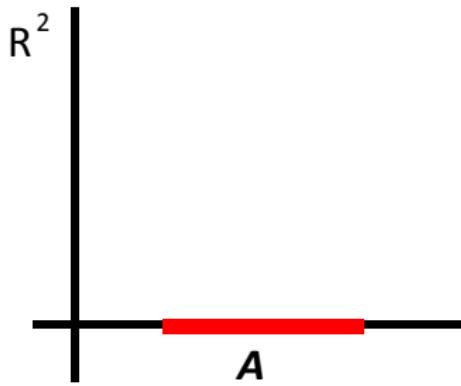


$X \setminus Y$	0	1	2	3	X
0	1/8	0	0	0	1/8
1	0	3/8	1/8	0	4/8
2	0	0	2/8	0	2/8
3	0	0	0	1/8	1/8
Y	1/8	3/8	3/8	1/8	1

Observación: $A \subset \mathbb{R}$, $\{X_1 \in A\} = \{\mathbf{X} \in A \times \mathbb{R} \times \cdots \times \mathbb{R}\}$.

Luego, $P(X_1 \in A) = P(\mathbf{X} \in A \times \mathbb{R} \times \cdots \times \mathbb{R})$.

Idea física: $n = 2$, $P(X \in A) = P((X, Y) \in A \times \mathbb{R})$



Relación entre la distribución conjunta y las marginales: La masa que la variable aleatoria X concentra en A es la masa que el vector aleatorio (X, Y) concentra en $A \times \mathbb{R}$.

2. Función de distribución conjunta

Dado $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, la **región suroeste** generada por \mathbf{x} es

$$S_{\mathbf{x}} = (-\infty, x_1] \times \cdots \times (-\infty, x_d].$$



Sea $\mathbf{X} = (X_1, \dots, X_d) : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}^d$ vector aleatorio. Se llama **función de distribución** de \mathbf{X} o **función de distribución conjunta** de las X_i -s a la función $F_{\mathbf{X}} : \mathbb{R}^d \longrightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \in S_{\mathbf{x}}) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

2. Función de distribución conjunta

Sea $\mathbf{X} = (X_1, \dots, X_k) : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^d$ vector aleatorio. Se llama **función de distribución** de \mathbf{X} o **función de distribución conjunta** de las X_i -s a la función $F_{\mathbf{X}} : \mathbb{R}^d \rightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \in S_{\mathbf{x}}) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Teorema: Propiedades básicas de la función de distribución

Sea \mathbf{X} vector aleatorio y F su función de distribución.

- ① F es no decreciente (en \mathbb{R}^d).
- ② F es continua por la derecha en cada variable.
- ③ $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_d) = 0$ ($i = 1, \dots, d$);
 $\lim_{x_1, \dots, x_d \rightarrow +\infty} F(x_1, \dots, x_d) = 1$.

Pregunta: ¿Qué significa que una función sea monótona (no decreciente) en \mathbb{R}^d ?

Funciones monótonas (no decrecientes) en \mathbb{R}^d

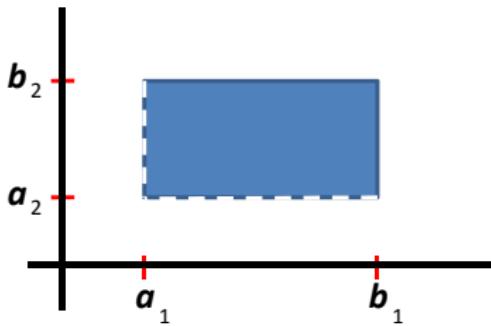
Notación: Si $G : \mathbb{R}^k \longrightarrow \mathbb{R}$ y $u \leq v$,

$$\Delta_i(u, v) = G(x_1, \dots, x_{i-1}, v, x_{i+1}, \dots, x_k) - G(x_1, \dots, x_{i-1}, u, x_{i+1}, \dots, x_k).$$

En $d = 1$, para $a \leq b$, $\Delta(a, b)F(x) = F(b) - F(a) \geq 0$.

En $d = 2$, para $a_1 \leq b_1$ y $a_2 \leq b_2$, se pide que

$$\begin{aligned}\Delta_1(a_1, b_1)\Delta_2(a_2, b_2)F(x, y) &= \Delta_1(a_1, b_1)(F(x, b_2) - F(x, a_2)) = \\ &= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0.\end{aligned}$$



En general, se pide que, para $a_i \leq b_i$ ($i = 1, \dots, d$)

$$\Delta_1(a_1, b_1)\Delta_2(a_2, b_2) \cdots \Delta_d(a_d, b_d)F(x_1, \dots, x_k) \geq 0.$$

2. Función de distribución conjunta

Sea $\mathbf{X} = (X_1, \dots, X_k) : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^d$ vector aleatorio.

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \in S_{\mathbf{x}}) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Teorema: Propiedades básicas de la función de distribución

Sea \mathbf{X} vector aleatorio y F su función de distribución.

- ① F es no decreciente (en \mathbb{R}^d).
- ② F es continua por la derecha en cada variable.
- ③ $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_d) = 0$ ($i = 1, \dots, d$);
 $\lim_{x_1, \dots, x_d \rightarrow +\infty} F(x_1, \dots, x_d) = 1$.

Teorema de unicidad

Sea $F : \mathbb{R}^d \rightarrow \mathbb{R}$ verificando ①, ② y ③ de arriba, entonces, existe $\mathbf{X} : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^d$ vector aleatorio tal que $F = F_{\mathbf{X}}$.

Además, el vector \mathbf{X} es único en distribución (es decir, la f.d. caracteriza la distribución de probabilidad de un v.a. de forma biunívoca).

2. Función de distribución conjunta

Sea $\mathbf{X} = (X_1, \dots, X_k) : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}^d$ vector aleatorio. Se llama **función de distribución** de \mathbf{X} o **función de distribución conjunta** de las X_i -s a la función $F_{\mathbf{X}} : \mathbb{R}^d \longrightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \in S_{\mathbf{x}}) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Cada una de las variables marginales, X_1, \dots, X_d , tiene una función de distribución (unidimensional):

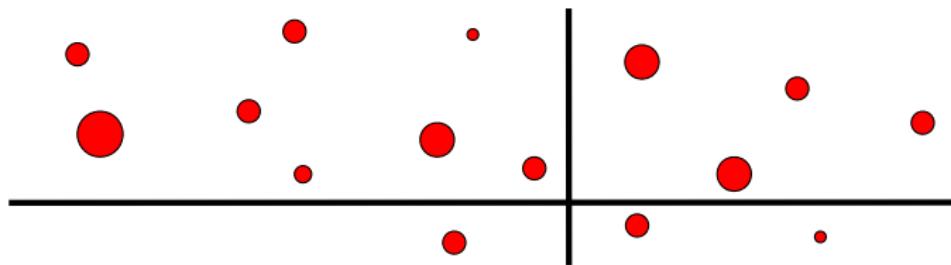
$$F_{X_i}(x_i) = P(X_i \leq x_i), \quad i = 1, \dots, d.$$

Las funciones F_1, \dots, F_d se llaman las **funciones de distribución marginales** de las variables X_1, \dots, X_d .

Pregunta: ¿Cuál es la relación entre $F_{\mathbf{X}}$ y F_{X_1}, \dots, F_{X_d} ?

$$F_{X_i}(x_i) = \lim_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_d).$$

Un vector aleatorio $\mathbf{X} : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^d$ se dice que tiene **distribución discreta** si existe un conjunto $S \subset \mathbb{R}^d$ contable (finito o numerable) tal que $P(\mathbf{X} \in S) = 1$ ($P(\mathbf{X} \in S^c) = 0$). Es decir, \mathbf{X} concentra su masa en S . El conjunto S se llama **soporte** de la distribución (de \mathbf{X}).



Nota: \mathbf{X} vector discreto con soporte $S \subset \mathbb{R}^d$, entonces

$$P(\mathbf{X} \in B) = \sum_{s \in S \cap B} P(\mathbf{X} = s), \quad B \subset \mathbb{R}^d.$$

En otras palabras, sólo nos interesa conocer cómo es la distribución de \mathbf{X} en un conjunto finito o infinito numerable S .

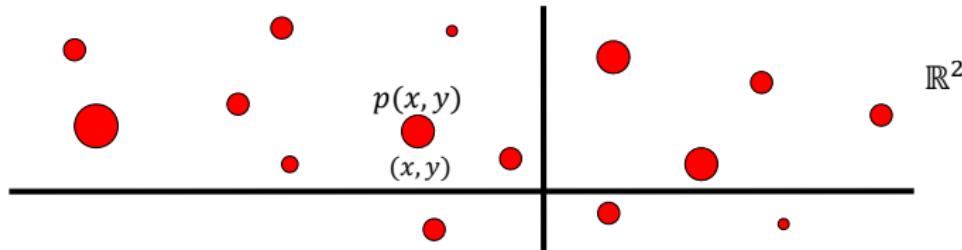
Sea $\mathbf{X} : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}^d$ un vector con distribución discreta y soporte S . La función:

$$p_{\mathbf{X}} : \mathbb{R}^d \longrightarrow \mathbb{R}$$

$$\mathbf{x} \longmapsto p_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$$

se denomina **función (de masa de) probabilidad conjunta** de \mathbf{X} .

Notación: Cuando no haya confusión denotaremos por p (en lugar de $p_{\mathbf{X}}$) la función de probabilidad del vector \mathbf{X} .



Si \mathbf{X} es un vector discreto con función de probabilidad p , tenemos:

- $p(\mathbf{x}) = 0$, si $\mathbf{x} \in S^c$ y $\sum_{\mathbf{x} \in S} p(\mathbf{x}) = 1$.
- $P(\mathbf{X} \in B) = \sum_{\mathbf{x} \in B \cap S} p(\mathbf{x})$ (suma finita o serie abs. conv.).

Un vector aleatorio $\mathbf{X} : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^d$ se dice **vector aleatorio discreto** si existe un conjunto $S \subset \mathbb{R}^d$ contable tal que $P(\mathbf{X} \in S) = 1$ ($P(\mathbf{X} \in S^c) = 0$). El conjunto S se llama **soporte** de la distribución (de \mathbf{X}).

Teorema: Vectores discretos

Sea $\mathbf{X} = (X_1, \dots, X_d) : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^d$ vector aleatorio.

\mathbf{X} vector discreto $\iff X_1, \dots, X_d$ variables discretas.

Si $\mathbf{X} = (X_1, \dots, X_d)$ vector discreto, las funciones de probabilidad de X_1, \dots, X_d , p_{X_1}, \dots, p_{X_d} , se denominan **funciones (de masa) de probabilidad marginales**.

Pregunta: ¿Qué relación hay entre $p_{\mathbf{X}}$ y p_{X_1}, \dots, p_{X_d} ?

$$p_{X_i}(x_i) = \sum_{j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_d} p_{\mathbf{X}}((x_{j_1}, \dots, x_{j_{i-1}}, x_i, x_{j_{i+1}}, \dots, x_{j_d})).$$

Una función $f : \mathbb{R}^d \rightarrow \mathbb{R}$ se llama **función de densidad (de probabilidad) conjunta** sobre \mathbb{R}^d si cumple:

- (a) $f(\mathbf{x}) \geq 0$, para todo $\mathbf{x} \in \mathbb{R}^d$.
- (b) f es integrable (Riemann).
- (c) $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_d) dt_1 \cdots dt_d = 1$.

Un vector aleatorio $\mathbf{X} = (X_1, \dots, X_d)$ con función de distribución F se dice que es **(absolutamente) continuo** si existe una función de densidad sobre \mathbb{R}^d , f , tal que para todo $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$

$$F(x_1, \dots, x_d) = \int_{S_x} f = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(t_1, \dots, t_d) dt_d \cdots dt_1. \quad (*)$$

De (*), y usando el teorema fundamental del cálculo, se tiene que en cada punto de continuidad (x_1, \dots, x_d) de f

$$\frac{\partial^d F(x_1, \dots, x_d)}{\partial x_1 \cdots \partial x_d} = f(x_1, \dots, x_d).$$

Cálculo de probabilidades con vectores continuos

Si $\mathbf{X} = (X_1, \dots, X_d)$ es un vector continuo con densidad f y $A \subset \mathbb{R}^d$ un conjunto “regular” para la integración, entonces

$$P(\mathbf{X} \in A) = \int\limits_A \cdots \int f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

Teorema: Vectores continuos

Si $\mathbf{X} = (X_1, \dots, X_d)$ es un vector aleatorio continuo, entonces X_1, \dots, X_d son variables continuas. Además, si \mathbf{X} tiene densidad f , la densidad de la variable marginal X_i ($i = 1, \dots, d$) es

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d.$$

f_i se llama **función de densidad marginal** de X_i ($i = 1, \dots, d$).

Atención: El recíproco del resultado anterior *no* es cierto.

Teorema: Vectores continuos

Si $\mathbf{X} = (X_1, \dots, X_d)$ es un vector aleatorio continuo, entonces X_1, \dots, X_d son variables continuas. Además, si \mathbf{X} tiene densidad f , la densidad de la variable marginal X_i ($i = 1, \dots, d$) es

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d.$$

f_i se llama **función de densidad marginal** de X_i ($i = 1, \dots, d$).

Ejercicio: Si $\mathbf{X} = (X, Y)$ vector aleatorio bidimensional con densidad conjunta $f(x, y)$. Las densidades de X e Y son:

$$f_1(x) =$$

$$f_2(y) =$$

Observación: Puede ocurrir que X_1, \dots, X_d variables aleatorias con densidad (es decir, cada una de ellas es absolutamente continua) y que el vector $\mathbf{X} = (X_1, \dots, X_d)$ no tenga densidad.

Ejemplo:

5. Ejemplos de distribuciones discretas multidimensionales

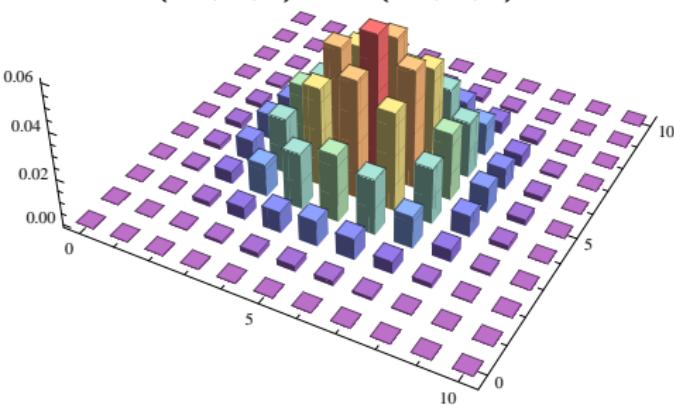
① Producto de binomiales ($n_1, n_2 \in \mathbb{N}, p_1, p_2 \in (0, 1)$)

$$P(X = k, Y = l) = \binom{n_1}{k} p_1^k (1 - p_1)^{n_1 - k} \binom{n_2}{l} p_2^l (1 - p_2)^{n_2 - l}.$$

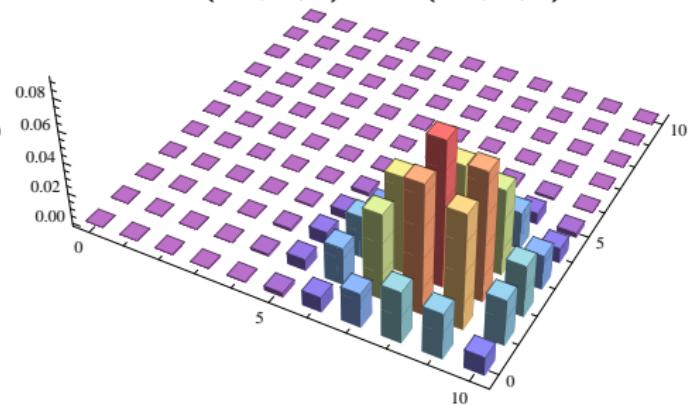
El soporte es $S = \{0, 1, \dots, n_1\} \times \{0, 1, \dots, n_2\}$.

Las marginales son $X \sim B(n_1, p_1)$ e $Y \sim B(n_2, p_2)$.

$$B(10, 0.5) \otimes B(10, 0.5)$$



$$B(10, 0.2) \otimes B(10, 0.8)$$



5. Ejemplos de distribuciones discretas multidimensionales

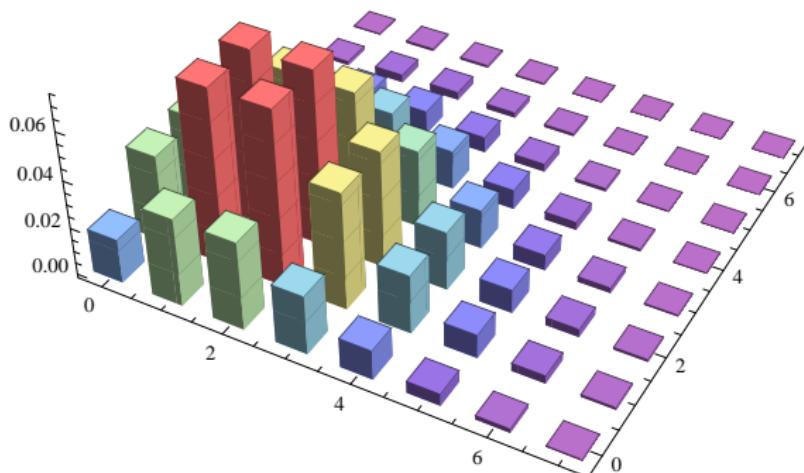
② Producto de Poisson de parámetros $\lambda, \mu > 0$

$$P(X = k, Y = l) = e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^l}{l!}, \quad k, l = 0, 1, \dots$$

El soporte es $S = \{0, 1, \dots\} \times \{0, 1, \dots\}$.

Las marginales son $X \sim P(\lambda)$ e $Y \sim P(\mu)$.

$$P(2) \otimes P(2)$$



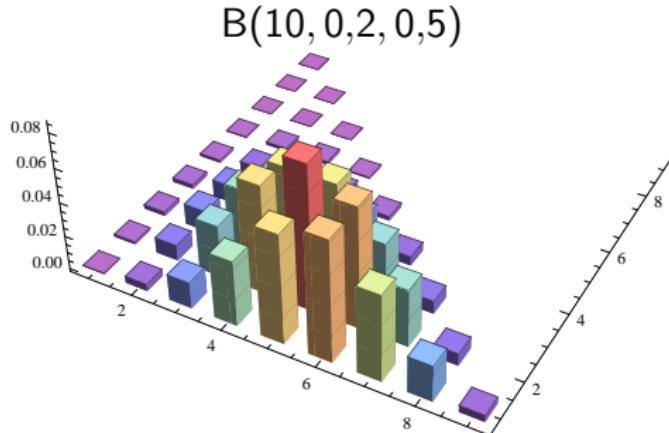
5. Ejemplos de distribuciones discretas multidimensionales

③ **Trinomial** ($n \in \mathbb{N}$, $p_1, p_2 \in (0, 1)$ con $p_1 + p_2 \leq 1$)

$$\begin{aligned} P(X = k, Y = l) &= \binom{n}{k} \binom{n-k}{l} p_1^k p_2^l (1 - p_1 - p_2)^{n-k-l} \\ &= \frac{n!}{k! l! (n-k-l)!} p_1^k p_2^l (1 - p_1 - p_2)^{n-k-l}. \end{aligned}$$

El soporte $S = \{(k, l) : k, l = 0, 1, \dots, n \text{ y } k + l \leq n\}$.

Las marginales son $X \sim B(n, p_1)$ e $Y \sim B(n, p_2)$.



6. Ejemplos de distribuciones continuas multidimensionales

Teorema: Producto de densidades univariadas

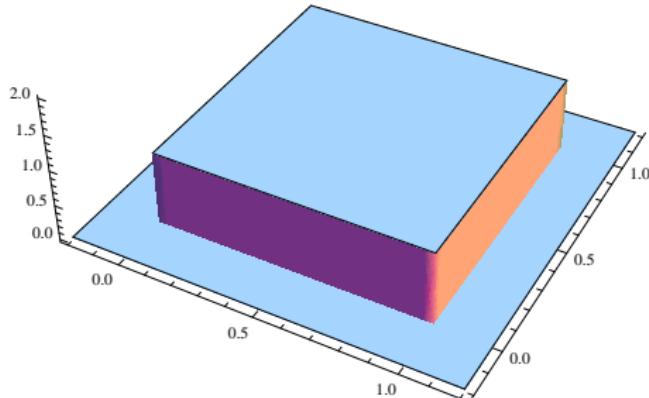
Sean f, g funciones de densidad sobre \mathbb{R} . La función

$$h(x, y) = f(x)g(y), \quad (x, y) \in \mathbb{R}^2$$

es una densidad en \mathbb{R}^2 cuyas densidades marginales son f y g .

① Uniforme en el rectángulo $(a, b) \times (c, d)$

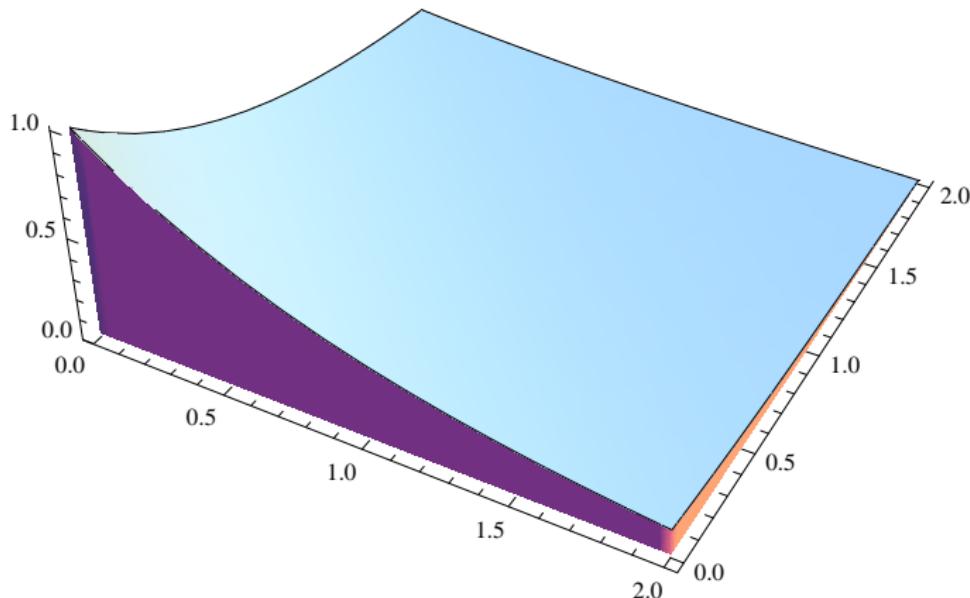
$$f(x, y) = \frac{1}{(b-a)(c-d)} \mathbf{1}_{(a,b) \times (c,d)}(x, y).$$



6. Ejemplos de distribuciones continuas multidimensionales

② Producto de exponentiales de parámetros $\lambda, \mu > 0$

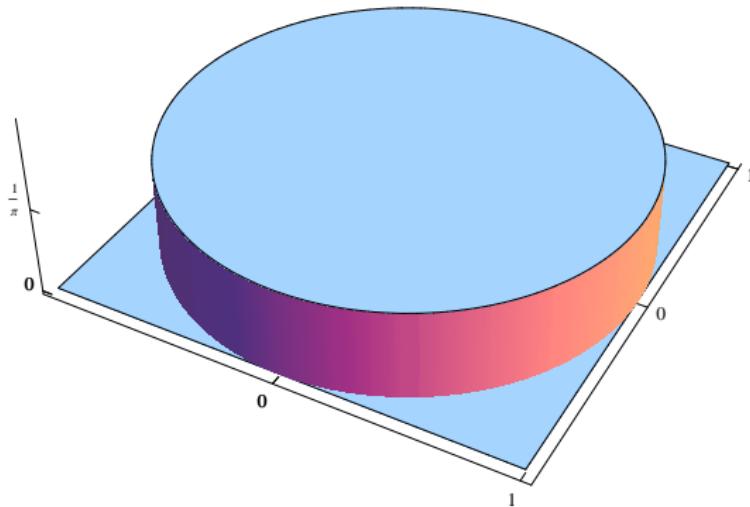
$$f(x, y) = \begin{cases} \lambda\mu e^{-(\lambda x + \mu y)}, & \text{si } x, y > 0, \\ 0, & \text{en otro caso.} \end{cases}$$



6. Ejemplos de distribuciones continuas multidimensionales

③ Uniforme en el círculo unidad

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

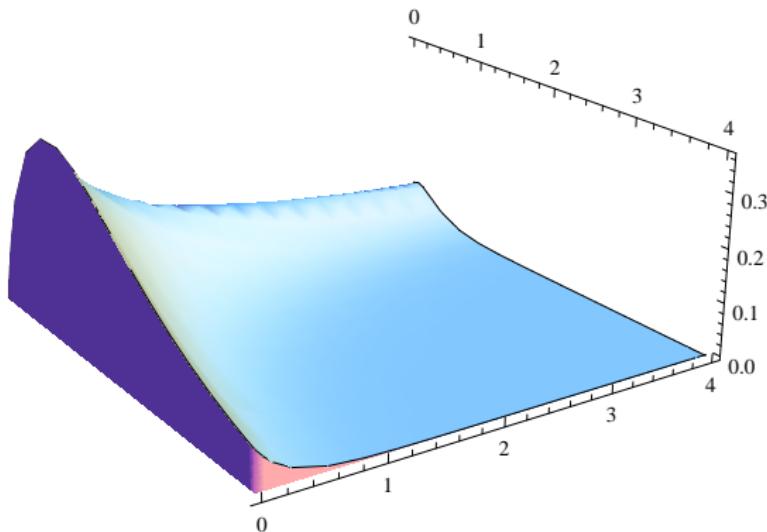


Ejercicio: Calcula las densidades marginales, $f_1(x)$ y $f_2(y)$.

6. Ejemplos de distribuciones continuas multidimensionales

④ (X, Y) con densidad:

$$f(x, y) = \begin{cases} xe^{-x(y+1)}, & \text{si } x, y > 0, \\ 0, & \text{en otro caso.} \end{cases}$$



Ejercicio: Calcula las densidades marginales, $f_1(x)$ y $f_2(y)$.

6. Ejemplos de distribuciones continuas multidimensionales

- ⑤ **Normal multivariante:** El vector aleatorio \mathbf{X} es **normal** (d -dimensional) con parámetros μ y Σ (notación: $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) si tiene densidad dada por:

$$f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-d/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}.$$

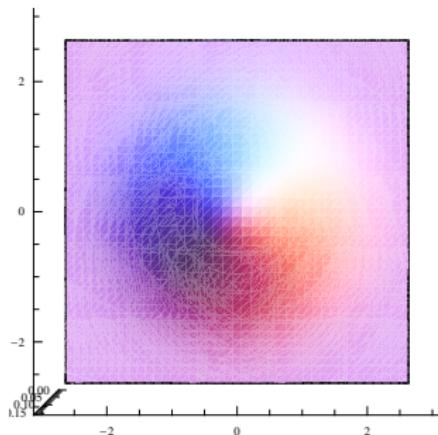
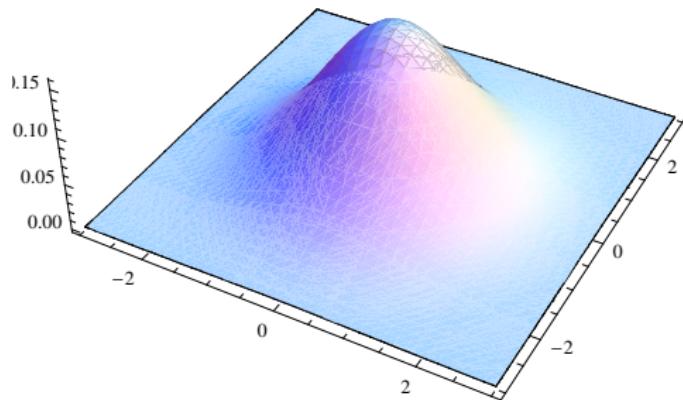
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{p2} & \cdots & \sigma_{dd} \end{bmatrix}.$$

Nota: Se puede comprobar que las variables marginales también son normales. De hecho, $X_i \sim N(\mu_i, \sigma_{ii})$.

6. Ejemplos de distribuciones continuas multidimensionales

Ejemplos de densidades normales en dimensión 2

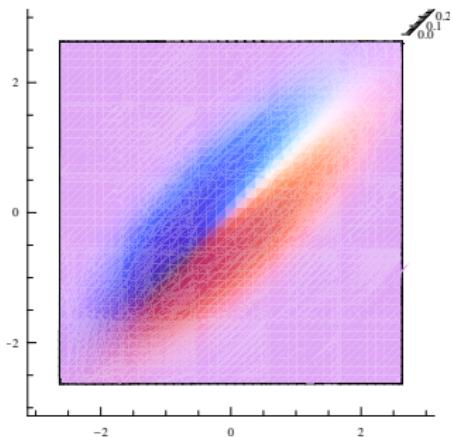
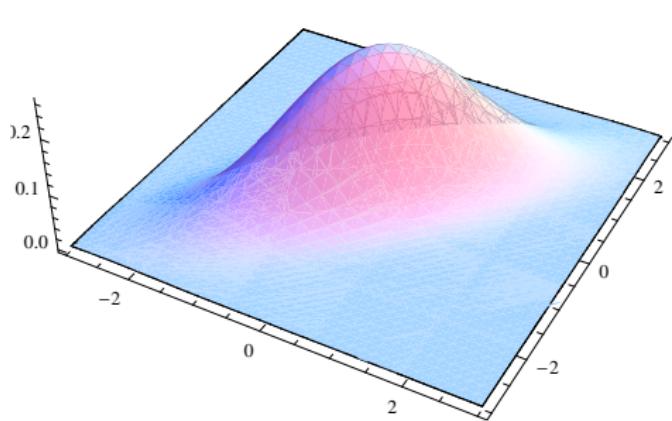
Densidad de $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = (0, 0)'$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.



6. Ejemplos de distribuciones continuas multidimensionales

Ejemplos de densidades normales en dimensión 2

Densidad de $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = (0, 0)'$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0,8 \\ 0,8 & 1 \end{pmatrix}$.



6. Ejemplos de distribuciones continuas multidimensionales

Nota: En dimensión 2, la densidad de la normal (bivariante) se suele expresar también como

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp\left[\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)\right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)} Q(x, y)\right], \end{aligned}$$

donde Q es la forma cuadrática

$$Q(x, y) = \frac{1}{1-\rho^2} \left(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}\right) \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \frac{x-\mu_1}{\sigma_1} \\ \frac{y-\mu_2}{\sigma_2} \end{pmatrix}.$$

Los parámetros del vector son $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ y $\rho \in (-1, 1)$.

7. Distribuciones condicionadas

CASO DISCRETO: Sea (X, Y) un vector discreto y $x_0 \in \mathbb{R}$ un valor *fijo* con $P(X = x_0) > 0$. La variable Y **condicionada** a $X = x_0$, $Y|X = x_0$, es la variable con distribución de probabilidad

$$p_{Y|X=x_0}(y) = P(Y = y|X = x_0) = \frac{P(X = x_0, Y = y)}{P(X = x_0)}, \quad y \in \mathbb{R}.$$

De forma análoga se define $X|Y = y_0$, para aquellos y_0 tales que $P(Y = y_0) > 0$.

CASO CONTINUO: Sea (X, Y) un vector aleatorio continuo con densidad conjunta f y $x_0 \in \mathbb{R}$ un valor *fijo* con $f_1(x_0) > 0$. La variable Y **condicionada** a $X = x_0$, $Y|X = x_0$, es la variable con densidad de probabilidad dada por

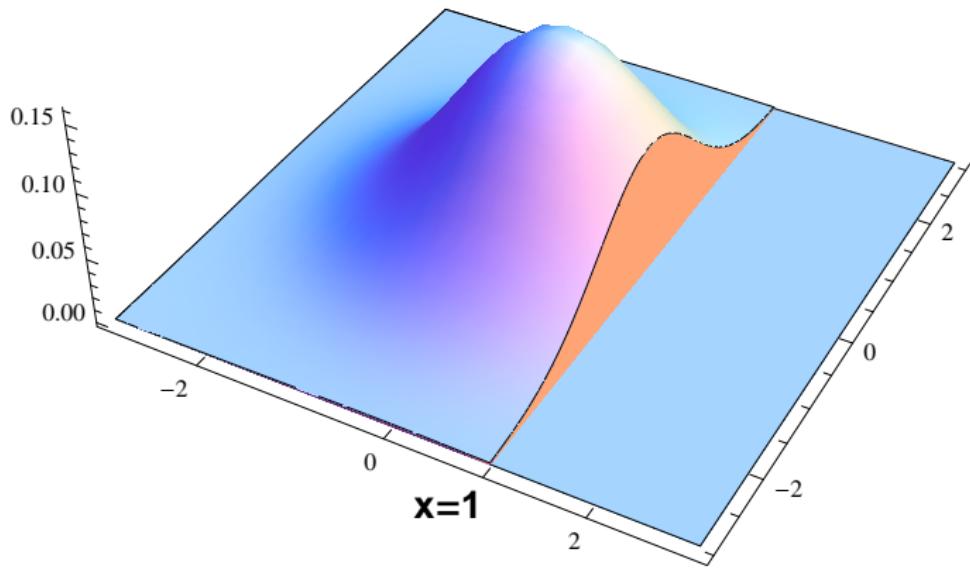
$$f_{Y|X=x_0}(y) = \frac{f(x_0, y)}{f_1(x_0)}, \quad y \in \mathbb{R}.$$

De forma análoga se define la densidad de $X|Y = y_0$, para aquellos y_0 tales que $f_2(y_0) > 0$.

7. Distribuciones condicionadas

Ejemplo: $\mathbf{X} = (X, Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = (0, 0)'$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

La densidad de $Y|X = 1$ es también normal $N(0, 1)$.



Ejercicio: (X, Y) vector aleatorio con densidad conjunta

$$f(x, y) = \begin{cases} xe^{-x(y+1)}, & \text{si } x, y > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

- (a) Para $a > 0$, calcula la densidad de la variable condicionada $Y|X = a$.
- (b) Para $a > 0$, calcula la densidad de la variable condicionada $X|Y = a$.

CASO DISCRETO: Dado un vector aleatorio discreto $\mathbf{X} = (X_1, \dots, X_d)$, se dice que las variables X_1, \dots, X_d son **independientes** si para cualesquiera $x_1, \dots, x_d \in \mathbb{R}$,

$$P(X_1 = x_1, \dots, X_d = x_d) = P(X_1 = x_1) \cdots P(X_d = x_d).$$

Es decir, la función de probabilidad conjunta de \mathbf{X} es el producto de las funciones de probabilidad marginales.

Observación: Si X_1, \dots, X_d son independientes, entonces para todo $x_1, \dots, x_d \in \mathbb{R}$, los d sucesos $\{X_1 = x_1\}, \dots, \{X_d = x_d\}$ son (mutuamente) independientes.

CASO CONTINUO: Dado un vector aleatorio continuo $\mathbf{X} = (X_1, \dots, X_d)$ con densidad conjunta f , se dice que las variables X_1, \dots, X_d son **independientes** si para cualesquiera $x_1, \dots, x_d \in \mathbb{R}$,

$$f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d).$$

Es decir, la función de densidad conjunta de \mathbf{X} es el producto de las funciones de densidad marginales.

Teorema: Caracterización mediante funciones de distribución

Sea $\mathbf{X} = (X_1, \dots, X_d)$ un vector (discreto o continuo) con función de distribución conjunta F y marginales F_1, \dots, F_d .

X_1, \dots, X_d independientes $\iff F(x_1, \dots, x_d) = F_1(x_1) \cdots F_d(x_d)$.

(La función de distribución conjunta es el producto de las funciones de distribución marginales.)

Distribución de la suma de variables: Dado un vector $\mathbf{X} = (X_1, \dots, X_d)$ con distribución de probabilidad conocida, $P_{\mathbf{X}}$, queremos calcular la distribución de la suma $S = X_1 + \dots + X_d$, P_S .

Sea (X, Y) un vector con densidad $h(x, y)$. La densidad de la variable $S = X + Y$ es

$$g_1(u) = \int_{-\infty}^{\infty} h(u - v, v) dv.$$

En particular, si X, Y son independientes con densidades $f_1(x)$ y $f_2(y)$, entonces la densidad de $X + Y$ es

$$g_1(u) = \int_{-\infty}^{\infty} f_1(u - v)f_2(v) dv = f_1 * f_2(u).$$

($f_1 * f_2$ es la **convolución** de f_1 con f_2 .)

9. Transformaciones de vectores aleatorios

Aplicación: Suma de variables normales independientes

- (a) Sean $X \sim N(0, \sigma)$, $Y \sim N(0, \tau)$ independientes. Mostrar que

$$X + Y \sim N\left(0, \sqrt{\sigma^2 + \tau^2}\right).$$

- (b) Sean $X \sim N(a, \sigma)$, $Y \sim N(b, \tau)$ independientes. Mostrar que

$$X + Y \sim N\left(a + b, \sqrt{\sigma^2 + \tau^2}\right).$$

- (c) Sean X_1, \dots, X_n variables independientes con $X_i \sim N(\mu_i, \sigma_i)$ ($i = 1, \dots, n$). Mostrar que

$$X_1 + \dots + X_n \sim N\left(\mu_1 + \dots + \mu_n, \sqrt{\sigma_1^2 + \dots + \sigma_n^2}\right).$$

- (d) Sean X_1, \dots, X_n variables independientes con $X_i \sim N(\mu_i, \sigma_i)$ ($i = 1, \dots, n$) y a_1, \dots, a_n constantes. Mostrar que

$$X_1 + \dots + X_n \sim N\left(a_1\mu_1 + \dots + a_n\mu_n, \sqrt{a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2}\right).$$

Caso discreto

Ejemplo 1: Sean $X \sim B(n_1, p)$, $Y \sim B(n_2, p)$ independientes. Mostrar que $X + Y \sim B(n_1 + n_2, p)$.

En general, si $X_i \sim B(n_i, p)$ ($i = 1, \dots, k$) independientes, entonces $X_1 + \dots + X_k \sim B(n_1 + \dots + n_k, p)$.

Ejemplo 2: Sean $X \sim P(\lambda)$, $Y \sim P(\mu)$ independientes. Mostrar que $X + Y \sim P(\lambda + \mu)$.

En general, si $X_i \sim P(\lambda_i)$ ($i = 1, \dots, n$) independientes, entonces $X_1 + \dots + X_n \sim P(\lambda_1 + \dots + \lambda_n)$.