# Distribuciones multivariantes

Mathias Bourel

IMERL - Facultad de Ingeniería, Universidad de la República, Uruguay

23 de abril de 2019

# Plan

# Plan

# Distribución conjunta de una variable aleatoria vectorial

Sea $\mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$ una VA vectorial.
Función de distribución:

$$F_{\mathbf{x}}(x^0) = \mathbb{P}(\mathbf{x} \leq x^0) = \mathbb{P}(x_1 \leq x_1^0, x_2 \leq x_2^0, \ldots, x_p \leq x_p^0)$$

- Si $\mathbf{x}$ es discreta, entonces $p(x^0) = \mathbb{P}(\mathbf{x} = x^0) = \mathbb{P}(x_1 = x_1^0, x_2 = x_2^0, \ldots, x_p = x_p^0)$
- Decimos que $\mathbf{x}$ es absolutamente continua, si existe una función $f : \mathbb{R}^P \rightarrow \mathbb{R}$ no negativa con $\int_{\mathbb{R}^p} f(\mathbf{x})\, d\mathbf{x} = 1$ tal que

$$F_{\mathbf{x}}(x^0) = \int_{-\infty}^{x^0} f(\mathbf{x})\, d\mathbf{x} = \int_{-\infty}^{x_1^0} \int_{-\infty}^{x_2^0} \cdots \int_{-\infty}^{x_p^0} f(x_1, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

Si $x$ es escalar y absolutamente continua, entonces
$p(x^0) = \mathbb{P}\left(x \in [x^0 - \frac{\Delta x}{2}, x^0 + \frac{\Delta x}{2}]\right) = \int_{x^0 - \frac{\Delta x}{2}}^{x^0 + \frac{\Delta x}{2}} f(t)\, dt \approx f(x^0)\Delta x$

En general si $\mathbf{x}$ es vectorial $p(x^0) = f(x^0)\Delta \mathbf{x}$, siendo $\Delta x$ el elemento de volumen.

# Distribuciones marginales y condicionadas

DISTRIBUCIONES MARGINALES:

- $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ con distribución conjunta $f_{x_1,x_2}$ entonces

$$f_{x_1}(x_1) = \int_{\mathbb{R}} f_{x_1,x_2}(x_1, x_2)\, dx_2 \qquad f_{x_2}(x_2) = \int_{\mathbb{R}} f_{x_1,x_2}(x_1, x_2)\, dx_1$$

  y con abuso de notación:

$$f(x_1) = \int_{\mathbb{R}} f(x_1, x_2)\, dx_2 \qquad f(x_2) = \int_{\mathbb{R}} f(x_1, x_2)\, dx_1$$

- $\int_{\mathbb{R}} f_{x_1}(x_1)\, dx_1 = \int_{\mathbb{R}} f_{x_2}(x_2)\, dx_2 = 1$

# Distribuciones marginales y condicionadas

DISTRIBUCIONES CONDICIONADAS:

- Sea el vector aleatorio $X = (\mathbf{x_1}, \mathbf{x_2}) \in \mathbb{R}^{p\times 2}$.
  Definimos la distribución condicionada de $\mathbf{x_1}$ para un valor de $\mathbf{x_2} = \mathbf{x_2^0}$ como

$$f(\mathbf{x_1}|\mathbf{x_2} = \mathbf{x_2^0}) = \frac{f(\mathbf{x_1}, \mathbf{x_2^0})}{f(\mathbf{x_2^0})} \quad \text{suponiendo que } f(\mathbf{x_2^0}) \neq 0$$

  Esto es consistente con el concepto de probabilidad condicionada, pues, suponiendo que las variables $\mathbf{x_1}$ y $\mathbf{x_2}$ son escalares

$$\underbrace{f(x_1|x_2^0)\Delta x_1}_{\mathbb{P}(x_1|x_2^0)} = \underbrace{\frac{f(x_1, x_2^0)\Delta x_1 \Delta x_2}{f(x_2^0)\Delta x_2}}_{\frac{\mathbb{P}(x_1, x_2^0)}{\mathbb{P}(x_2^0)}}$$

  Entonces

$$f(\mathbf{x_1}, \mathbf{x_2}) = f(\mathbf{x_1}|\mathbf{x_2}) f(\mathbf{x_2}) \qquad f(\mathbf{x_1}, \mathbf{x_2}) = f(\mathbf{x_2}|\mathbf{x_1}) f(\mathbf{x_1})$$

# Distribuciones marginales y condicionadas

La distribución marginal de $\mathbf{x}_2$ se puede calcular como

$$f(\mathbf{x}_2) = \int f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1) d\,\mathbf{x}_1$$

Observar que si multiplicamos por $\Delta\mathbf{x}_2$ esto se puede interpretar como

$$f(\mathbf{x}_2)\Delta\mathbf{x}_2 = \int f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1) d\,\mathbf{x}_1 \Delta\mathbf{x}_2$$

$$p(\mathbf{x}_2) = \sum \underbrace{f(\mathbf{x}_2|\mathbf{x}_1)\Delta\mathbf{x}_2}_{p(\mathbf{x}_2|\mathbf{x}_1)} \underbrace{f(\mathbf{x}_1)\Delta\mathbf{x}_1}_{p(\mathbf{x}_1)}$$

FORMULA DE BAYES:

$$f(\mathbf{x}_1|\mathbf{x}_2) = \frac{f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1)}{f(\mathbf{x}_2)} = \frac{f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1)}{\int f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1) d\,\mathbf{x}_1}$$

# Ejemplo

$x_1$ votar a algun candidato $c_1, c_2, c_3, c_4$ y $x_2$ nivel de ingreso A (alto), M (medio), B (bajo). Presentamos la distribución conjunta de votos:

| | A | M | B |
|---|---|---|---|
| $c_1$ | .1 | .05 | .01 |
| $c_2$ | .05 | .20 | .04 |
| $c_3$ | .04 | .25 | .07 |
| $c_4$ | .01 | .1 | .08 |

Distribuciones marginales:

| | A | M | B | votos |
|---|---|---|---|---|
| $c_1$ | .1 | .05 | .01 | 0.16 |
| $c_2$ | .05 | .20 | .04 | 0.29 |
| $c_3$ | .04 | .25 | .07 | 0.36 |
| $c_4$ | .01 | .1 | .08 | 0.19 |
| ingresos | 0.2 | 0.6 | 0.2 | |

# Ejemplo

Distribución condicionada de los votos por personas con nivel de ingreso B:

| | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|
| B | $\frac{0{,}01}{0{,}2} = 0{,}05$ | $\frac{0{,}04}{0{,}2} = 0{,}2$ | $\frac{0{,}07}{0{,}2} = 0{,}35$ | $\frac{0{,}08}{0{,}2} = 0{,}4$ |

Distribución condicionada de los ingresos por votantes del candidato $c_4$:

| | A | M | B |
|---|---|---|---|
| $c_4$ | $\frac{0{,}01}{0{,}19} = 0{,}0526$ | $\frac{0{,}1}{0{,}19} = 0{,}5263$ | $\frac{0{,}08}{0{,}19} = 0,4211$ |

# Independencia de vectores aleatorios

$\mathbf{x_1}$ y $\mathbf{x_2}$ son **independientes** si

$$f(\mathbf{x_2}|\mathbf{x_1}) = f(\mathbf{x_2})$$

lo cual equivale a

$$f(\mathbf{x_1}, \mathbf{x_2}) = f(\mathbf{x_1})f(\mathbf{x_2})$$

Notación: $\mathbf{x_1} \perp \mathbf{x_2}$

Observación: si $\mathbf{x_1} \perp \mathbf{x_2}$ entonces $g_1(\mathbf{x_1}) = \mathbf{y_1} \perp \mathbf{y_2} = g_2(\mathbf{x_2})$

# Vector de medias

Sea $\mathbf{x} = (x_1, \ldots, x_p)$ una VA en $\mathbb{R}^p$. El vector de medias es

$$\mathbb{E}(\mathbf{x}) = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \mathbb{E}(x_1) \\ \vdots \\ \mathbb{E}(x_p) \end{pmatrix}$$

Si $\mathbf{x}$ es continua entonces notamos $\mathbb{E}(\mathbf{x}) = \int \mathbf{x} f(\mathbf{x}), d\mathbf{x}$ donde $\mathbb{E}(x_i) = \int x f_{x_i}(x)\, dx\, \forall\, i$

## Proposición 1

- $\mathbb{E}(A\mathbf{x} + \mathbf{b}) = A\mathbb{E}(\mathbf{x}) + \mathbf{b}$ *siendo $A$ una matriz y $\mathbf{b}$ un vector.*
- *Si $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2})$ y $a, b \in \mathbb{R}$ entonces $\mathbb{E}(a\mathbf{x_1} + b\mathbf{x_2}) = a\mathbb{E}(\mathbf{x_1}) + b\mathbb{E}(\mathbf{x_2})$*

# Matriz de varianzas y covarianzas

Sea $\mathbf{x} = (x_1, \ldots, x_p)$ una VA en $\mathbb{R}^p$. Su matriz de varianzas y covarianzas es la matriz cuadarada

$$Var(\mathbf{x}) = V_x = \mathbb{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'] \in \mathcal{M}_{p \times p}$$

$$V_x = \begin{pmatrix} \sigma_1^2 & s_{12} & \ldots & s_{1p} \\ s_{12} & \sigma_2^2 & \ldots & s_{2p} \\ \vdots & & \ddots & \vdots \\ s_{1p} & s_{2p} & \ldots & \sigma_p^2 \end{pmatrix}$$

siendo

$$\sigma_i^2 = Var(x_i), \;\; s_{ij} = Cov(x_i, x_j) = \mathbb{E}(x_i x_j) - \mathbb{E}(x_i)\mathbb{E}(x_j)$$

# Matriz de varianzas y covarianzas

Propiedades:

1. $V_x$ es simétrica (es claro)
2. $V_x$ es semidefinida positiva, es decir para todo $\mathbf{w} \in \mathbb{R}^p$ se tiene que $\mathbf{w}' V_x \mathbf{w} \geq 0$.
   En efecto, sea $\mathbf{w} \in \mathbb{R}^p$ y defino $y = (\mathbf{x} - \mu)'\mathbf{w} \in \mathbb{R}$. Entonces
   $$\mathbb{E}(y) = \mathbb{E}((\mathbf{x} - \mu)'\mathbf{w}) = \mathbb{E}((\mathbf{x} - \mu))'\mathbf{w} = 0$$
   Por lo tanto
   $$Var(y) = \mathbb{E}(y^2) = \mathbf{w}'\mathbb{E}((\mathbf{x} - \mu)(\mathbf{x} - \mu)')\mathbf{w} = \mathbf{w}' V_x \mathbf{w} \geq 0$$

# Transformación de vectores aleatorios

Sea $\mathbf{x} = (x_1, \ldots, x_p)$ un vector aleatorio de $\mathbb{R}^p$ con densidad $f_x(\mathbf{x})$ y sea otro vector aleatorio $\mathbf{y} = (y_1, \ldots, y_p) \in \mathbb{R}^p$ definido por

$$\begin{cases} y_1 = g_1(x_1, \ldots, x_p) \\ y_2 = g_2(x_1, \ldots, x_p) \\ \qquad \vdots \\ y_p = g_p(x_1, \ldots, x_p) \end{cases}$$

donde suponemos que existen las funciones inversas $x_1 = h_1(y_1, \ldots, y_p), \ldots, x_p = h_p(y_1, \ldots, y_p)$, siendo $g_1, \ldots, g_p, h_1, \ldots, h_p$ diferenciables.

Entonces puede demostrarse (regla del Jacobiano) que:

$$f_y(\mathbf{y}) = f_x(\mathbf{x}) \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \ldots & \ldots & \frac{\partial x_1}{\partial y_p} \\ \vdots & & & \vdots \\ \frac{\partial x_p}{\partial y_1} & \ldots & \ldots & \frac{\partial x_p}{\partial y_p} \end{pmatrix} \right|$$

Caso particular. Supongamos que $\mathbf{y} = A\mathbf{x}$ siendo $A \in \mathcal{M}_{p \times p}$ invertible. Entonces

$$f_y(\mathbf{y}) = f_x(A^{-1}\mathbf{y}) |\det(A^{-1})|$$

# Esperanza y varianza de transformaciones lineales

Supongamos que $\mathbf{y} = A\mathbf{x}$ siendo $A \in \mathcal{M}_{p\times p}$. Entonces

## Proposición 2

1. $\mu_y = A\mu_x$
2. $V_y = AV_xA'$

## Demostración.

1. Ya lo vimos.
2. $V_y = \mathbb{E}((\mathbf{y} - \mu_y)(\mathbf{y} - \mu_y)')$
   $= \mathbb{E}(A(\mathbf{x} - \mu_x)(\mathbf{x} - \mu_x)'A') = A\mathbb{E}((\mathbf{x} - \mu_x)(\mathbf{x} - \mu_x)')A' = AV_xA'$

$\square$

# Esperanza condicionada

La esperanza de $\mathbf{x_1}$ condicionada a $\mathbf{x_2}$ es:

$$\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}) = \int \mathbf{x_1} f(\mathbf{x_1}|\mathbf{x_2})\, d\mathbf{x_1}$$

y es una función de $\mathbf{x_2}$.
Si $\mathbf{x_2}$ es un valor fijo entonces $\mathbb{E}(\mathbf{x_1}|\mathbf{x_2})$ es constante.
Si $\mathbf{x_2}$ es un variable aleatoria entonces $\mathbb{E}(\mathbf{x_1}|\mathbf{x_2})$ es una variable aleatoria.

# Esperanza condicionada

## Proposición 3

*Se cumple que*

$$\mathbb{E}(x_1) = \mathbb{E}\big(\mathbb{E}(x_1|x_2)\big)$$

*y la esperanza de la media condicionada es la esperanza marginal*

Demostración:

$$\mathbb{E}(x_1) = \int x_1 f(x_1) dx_1 = \int x_1 \left( \int f(x_1|x_2) f(x_2) dx_2 \right) dx_1$$

$$= \int \int x_1 f(x_1|x_2) f(x_2) dx_1 dx_2 = \int f(x_2) \left( \int x_1 f(x_1|x_2) dx_1 \right) dx_2$$

$$= \int f(x_2) \mathbb{E}(x_1|x_2) dx_2 = \mathbb{E}\big(\mathbb{E}(x_1|x_2)\big)$$

# Varianza condicionada

$$\mathrm{Var}(\mathbf{x_1}|\mathbf{x_2}) = \mathbb{E}[(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))^2|\mathbf{x_2}]$$

## Propiedad

$$\mathrm{Var}(\mathbf{x_1}|\mathbf{x_2}) = \mathbb{E}({\mathbf{x_1}}^2|\mathbf{x_2}) - (\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))^2$$

Lo verificamos para cada $y$:

$$\begin{aligned}\mathrm{Var}(\mathbf{x_1}|\mathbf{x_2}=y) =& \mathbb{E}[(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{x_2}=y))^2|\mathbf{x_2}=y) \\ =& \mathbb{E}[{\mathbf{x_1}}^2 + (\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}=y))^2 - 2\mathbf{x_1}\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}=y)|\mathbf{x_2}=y)\end{aligned}$$

Como $\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}=y)$ es una constante cuando se considera la distribución condicionada a $\mathbf{x_2}=y$, el resultado lo deducimos de la linealidad de la esperanza condicionada.

## Descomposición de la varianza

$$\mathrm{Var}(\mathbf{x_1}) = \mathrm{Var}(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2})) + \mathbb{E}(\mathrm{Var}(\mathbf{x_1}|\mathbf{x_2}))$$

Se sabe que $\mathbb{E}(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2})) = \mathbb{E}(\mathbf{x_1})$ y $\mathbb{E}(\mathbb{E}({\mathbf{x_1}}^2|\mathbf{x_2})) = \mathbb{E}({\mathbf{x_1}}^2)$. Entonces:

$$\begin{aligned}\mathrm{Var}(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2})) + \mathbb{E}(\mathrm{Var}(\mathbf{x_1}|\mathbf{x_2})) =& \mathbb{E}[(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))^2] - [\mathbb{E}(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))]^2 + \mathbb{E}(\mathbb{E}({\mathbf{x_1}}^2|\mathbf{x_2})) - \mathbb{E}[(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))^2] \\ =& \mathbb{E}[(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))^2] - [\mathbb{E}(\mathbf{x_1})]^2 + \mathbb{E}({\mathbf{x_1}}^2) - \mathbb{E}[(\mathbb{E}(\mathbf{x_1}|\mathbf{x_2}))^2] = \mathrm{Var}(\mathbf{x_1}).\end{aligned}$$

## Descomposición de la varianza

En particular, si $x_1$ es una variable aleatoria real y $\mathbb{E}(x_1) = \mu_1$ entonces:

- $(A)$ : $\mathbb{E}\big((x_1 - \mu_1)^2\big) = \mathrm{Var}(x_1)$
- $(B)$ : como $\mu_1 = \mathbb{E}(\mathbb{E}(x_1|\mathbf{x_2}))$ entonces $\mathrm{Var}(\mathbb{E}(x_1|\mathbf{x_2})) = \mathbb{E}\big((\mathbb{E}(x_1|\mathbf{x_2}) - \mu_1)^2\big)$
- $(C)$ : como $\mathrm{Var}(x_1|\mathbf{x_2}) = \mathbb{E}((x_1 - \mathbb{E}(x_1|\mathbf{x_2}))^2|\mathbf{x_2})$ entonces $\mathbb{E}(\mathrm{Var}(x_1|\mathbf{x_2})) = \mathbb{E}\big((x_1 - \mathbb{E}(x_1|\mathbf{x_2}))^2\big)$

Por lo tanto de

$$\mathrm{Var}(x_1) = \mathbb{E}(\mathrm{Var}(x_1|\mathbf{x_2})) + \mathrm{Var}(\mathbb{E}(x_1|\mathbf{x_2}))$$

se deduce que:

$$\underbrace{\mathbb{E}\big((x_1 - \mu_1)^2\big)}_{(A)} = \underbrace{\mathbb{E}\big((\mathbb{E}(x_1|\mathbf{x_2}) - \mu_1)^2\big)}_{(B)} + \underbrace{\mathbb{E}\big((x_1 - \mathbb{E}(x_1|\mathbf{x_2}))^2\big)}_{(C)}$$

El segundo termino promedia las varianzas de las distribuciones condicionadas. El primer termino recoge las diferencias entre la media global $\mu_1$ y las medias condicionadas.

- Si $x_1 \perp \mathbf{x_2}$ entonces

$$\mathbb{E}(x_1|\mathbf{x_2}) = \int x_1 f(x_1|\mathbf{x_2})\, dx_1 = \int x_1 f(x_1)\, dx_1 = \mathbb{E}(x_1) = \mu_1$$

  y por lo tanto el segundo miembro $var(\mathbb{E}(x_1|\mathbf{x_2})) = 0$

- En modelos lineales univariantes, si $\overline{x}$ es la media global:

$$\frac{1}{n}\sum(x_i - \overline{x})^2 = \underbrace{\frac{1}{n}\sum(x_i - \widehat{x}_i)^2}_{(C)} + \underbrace{\frac{1}{n}\sum(\widehat{x}_i - \overline{x})^2}_{(B)}$$

# Plan

# Distribución binomial

1. Queremos clasificar en dos clases $A$ y $B$. Supongamos que $\mathbb{P}(A) = p(= cte)$. La variable aleatoria considerada es

$$x = \begin{cases} 1, & \text{si la observación pertenece a A} \\ 0, & \text{si la observación no pertenece a A} \end{cases}$$

   Entonces $x \sim Ber(p)$ con $\mathbb{P}(x = 1) = p$ y $\mathbb{P}(x = 0) = 1 - p$.

2. Supongamos que repetimos independientemente este experimento y tenemos $n$ observaciones $x_1, x_2, \ldots, x_n$. Defino $y = \sum_{i=1}^{n} x_i$ con $x_i \sim Ber(p)$. Se tiene que

$$\mathbb{P}(y = r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \quad \forall\, r = 0, \ldots, n$$

   Entonces $y \sim Bin(n, p)$

# Distribución multinomial

- Supongamos que ahora tenemos $G$ clases y $\mathbf{p} = (p_1, \dots, p_G)$ es un vector de probabilidad $(\sum_{g=1}^{G} p_g = 1)$. Definimos

$$x_j = \begin{cases} 1, & \text{si la observación pertenece a } G_j \\ 0, & \text{si la observación no pertenece a } G_j \end{cases} \quad \forall\, j = 1, \dots, G$$

Entonces $\mathbf{x} = (x_1, \dots, x_G)$ es de la forma $(0, \dots, 0, 1, 0, \dots, 0)$. Las componentes de esta variable aleatoria no son independientes pues $\sum_{j=1}^{G} x_j = 1$, su distribución se llama *multinomial puntual* y tiene como función de probabilidad

$$\mathbb{P}(x_1, \dots, x_G) = p_1^{x_1} \dots p_G^{x_G}$$

y la probabilidad de que la $j$-esima coordenada sea 1 es $p_j$.

# Distribución multinomial

- Sea $\mathbf{x_1}, \ldots, \mathbf{x_n}$ una muestra de $n$ vectores de esta VA multinomial puntual. Por lo tanto esto es equivalente a clasificar $n$ elementos de una muestra en $G$ clases. Llamamos *distribución multinomial* a la distribución del vector

$$\mathbf{y} = \sum_{i=1}^{n} \mathbf{x_i}$$

Las componentes del vector $\mathbf{y}$ corresponden a las frecuencias con que se observan cada clase en la muestra:

$$y_i \in \{0, 1, \ldots, n\} \qquad \sum_{i=1}^{G} y_i = n$$

$$\mathbb{P}(y_1 = n_1, \ldots, y_G = n_G) = \frac{n!}{n_1! \ldots n_G!} p_1^{n_1} \ldots p_G^{n_G} \qquad n = \sum_{i=1}^{G} n_i$$

# Distribución multinomial

- Sea $\mathbf{x_1}, \dots, \mathbf{x_n}$ una muestra de $n$ vectores de esta VA multinomial puntual. Por lo tanto esto es equivalente a clasificar $n$ elementos de una muestra en $G$ clases. Llamamos *distribución multinomial* a la distribución del vector

$$\mathbf{y} = \sum_{i=1}^{n} \mathbf{x_i}$$

Las componentes del vector $\mathbf{y}$ corresponden a las frecuencias con que se observan cada clase en la muestra:

$$y_i \in \{0, 1, \dots, n\} \qquad \sum_{i=1}^{G} y_i = n$$

$$\mathbb{P}(y_1 = n_1, \dots, y_G = n_G) = \frac{n!}{n_1! \dots n_G!} p_1^{n_1} \dots p_G^{n_G} \qquad n = \sum_{i=1}^{G} n_i$$

Si $\mathbf{y}$ tiene distribución multinomial notamos:

$$\mathbf{y} \sim Mult(n, \mathbf{p})$$

En este caso (ejercicio):

1. $\mathbb{E}(\mathbf{y}) = n\mathbf{p} = \mu_y$
2. $Var(\mathbf{y}) = n(diag(\mathbf{p}) - \mathbf{p}\mathbf{p}') = diag(\mu_y) - \frac{1}{n}\mu_y \mu_y'$

# Distribución multinomial (ejemplo)

10 personas votan por 4 candidatos con probabilidades

$$\mathbb{P}(1) = \frac{1}{2}, \quad \mathbb{P}(2) = \mathbb{P}(3) = \frac{1}{8}, \quad \mathbb{P}(4) = \frac{1}{4}$$

Calcular la probabilidad que $(y_1, y_2, y_3, y_4) = (6, 1, 2, 1)$:

# Distribución multinomial (ejemplo)

10 personas votan por 4 candidatos con probabilidades

$$\mathbb{P}(1) = \frac{1}{2}, \quad \mathbb{P}(2) = \mathbb{P}(3) = \frac{1}{8}, \quad \mathbb{P}(4) = \frac{1}{4}$$

Calcular la probabilidad que $(y_1, y_2, y_3, y_4) = (6, 1, 2, 1)$:

$$\mathbb{P}(y_1 = 6, y_2 = 1, y_3 = 2, y_4 = 1) = \frac{10!}{6!1!2!1!} \left(\frac{1}{2}\right)^6 \left(\frac{1}{8}\right)^1 \left(\frac{1}{8}\right)^2 \left(\frac{1}{4}\right)^1$$

# Plan

# Normal univariada

Recordamos:

1. la densidad de una normal univariada típica

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

   y escribimos $x \sim N(0,1)$

# Normal univariada

Recordamos:

1. la densidad de una normal univariada típica
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
y escribimos $x \sim N(0,1)$
2. la densidad de una normal univariada
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
y escribimos $x \sim N(\mu, \sigma^2)$

# Normal univariada

Recordamos:

1. la densidad de una normal univariada típica

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

   y escribimos $x \sim N(0,1)$

2. la densidad de una normal univariada

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

   y escribimos $x \sim N(\mu, \sigma^2)$

Observaciones:

- Si $\sigma$ es grande el pico de la gaussiana es chico y si $\sigma$ es chico el pico de la gaussiana es grande.
- Si $x \sim N(\mu, \sigma^2)$ entonces $z = \frac{x-\mu}{\sigma} \sim N(0,1)$.

# Normal univariada

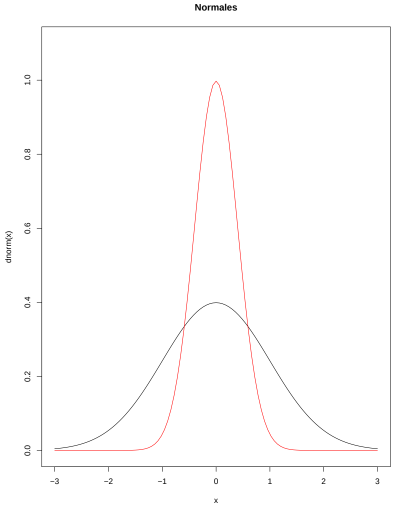
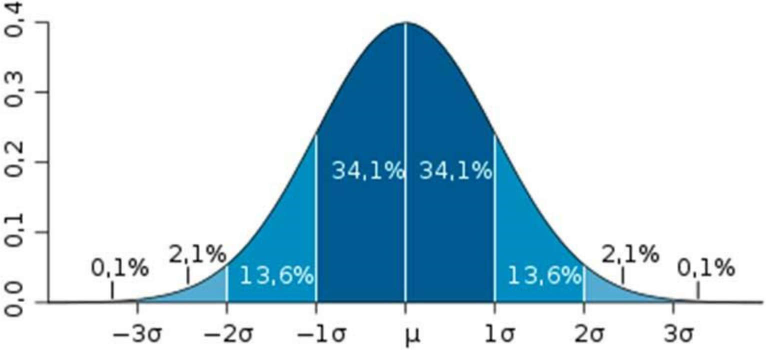



Figura: En negro $N(0,1)$ y en rojo $N(0,0.4)$

# Normal univariada




$$\mathbb{P}(\mu-\sigma < X < \mu+\sigma) = \mathbb{P}\left(-1 < \frac{X-\mu}{\sigma} < 1\right) = \Phi(1)-\Phi(-1) = 2\Phi(1)-1 \approx 2\times 0{,}84-1 \approx 0,682$$

# Normal multivariada

Decimos que $\mathbf{x} = (x_1, \ldots, x_d)$ es **gaussiana típica** en $\mathbb{R}^d$ si tiene densidad conjunta

$$f_x(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}||\mathbf{x}||^2}$$

Notación:

$$\mathbf{x} \sim N(0_{\mathbb{R}^d}, I_d)$$

# Normal multivariada

Decimos que $\mathbf{x} = (x_1, \ldots, x_d)$ es **gaussiana típica** en $\mathbb{R}^d$ si tiene densidad conjunta

$$f_x(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}||\mathbf{x}||^2}$$

Notación:

$$\mathbf{x} \sim N(0_{\mathbb{R}^d}, I_d)$$

Decimos que $\mathbf{y}$ es **gaussiana** en $\mathbb{R}^d$ o **normal multivariada** con media $\mu \in \mathbb{R}^d$ y matriz de covarianza $\Sigma = AA'$ siendo $A \in \mathcal{M}_{d\times d}$ si $\mathbf{y}$ tiene la misma distribución que $\mu + A\mathbf{x}$ con $\mathbf{x} \sim N(0_{\mathbb{R}^d}, I_d)$.

Notación:

$$\mathbf{y} \sim N(\mu, \Sigma)$$

# Normal multivariada: ejemplo

Sea $\mathbf{x} = (x_1, x_2) \sim N(0_{\mathbb{R}^2}, I_2)$

Si defino

$$\mathbf{y} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Entonces

$$\mathbf{y} \sim N\left( \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

# Normal bivariada

library(mvtnorm) library(MASS)

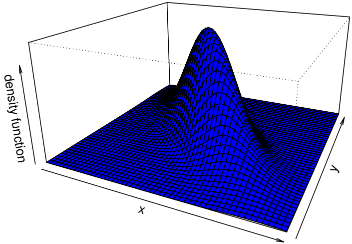



Figura: $N(\mu, \Sigma)$ con $\mu = (0,0)$ y $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

# Densidad de una normal $N(\mu, \Sigma)$

Si $\mathbf{y}$ es **gaussiana** en $\mathbb{R}^d$, $\mathbf{y} \sim N(\mu, \Sigma)$ con $\boldsymbol{\Sigma} = AA'$ e $\mathbf{y} = \mu + A\mathbf{x}$ entonces

$$\mathbf{y} = Q(\mathbf{x})$$

con $Q : \mathbb{R}^d \to \mathbb{R}^d$ definida por $Q(\mathbf{x}) = \mu + A\mathbf{x}$.

[1] Si $y = Ax + \mu$ entonces $f_y(y) = f_x(A^{-1}(y - \mu))|det(A)^{-1}|$

# Densidad de una normal $N(\mu, \Sigma)$

Si $\mathbf{y}$ es **gaussiana** en $\mathbb{R}^d$, $\mathbf{y} \sim N(\mu, \Sigma)$ con $\Sigma = AA'$ e $\mathbf{y} = \mu + A\mathbf{x}$ entonces

$$\mathbf{y} = Q(\mathbf{x})$$

con $Q : \mathbb{R}^d \to \mathbb{R}^d$ definida por $Q(\mathbf{x}) = \mu + A\mathbf{x}$.

Supongamos que $A$ es invertible, entonces $Q$ es biyectiva, $J_Q = \det(A)$ y por lo tanto, por el teorema del Jacobiano [1], $Y$ es absolutamente continua con densidad:

$$f_y(\mathbf{y}) = f_{Q(x)}(\mathbf{y}) = f_X(A^{-1}(\mathbf{y} - \mu))\frac{1}{\sqrt{\det(\Sigma)}} = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}||A^{-1}(\mathbf{y}-\mu)||^2}$$

$$= \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(A^{-1}(\mathbf{y}-\mu))'A^{-1}(\mathbf{y}-\mu)} = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{y}-\mu)'\Sigma^{-1}(\mathbf{y}-\mu)}$$

y por lo tanto

$$f_y(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}$$

[1] Si $y = Ax + \mu$ entonces $f_y(y) = f_x(A^{-1}(y - \mu))|det(A)^{-1}|$

# Normal multivariada (teoremas)

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir $\Sigma$ es una matriz diagonal).

# Normal multivariada (teoremas)

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir $\Sigma$ es una matriz diagonal).

    - Es bien conocido que si $x_1, \ldots, x_d$ son independientes entonces $x_1, \ldots, x_d$ están incorrelacionadas y esto implica que la matriz de varianzas del vector $\mathbf{x} = (x_1, \ldots, x_d)$ es diagonal.

# Normal multivariada (teoremas)

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir $\Sigma$ es una matriz diagonal).

   - Es bien conocido que si $x_1, \ldots, x_d$ son independientes entonces $x_1, \ldots, x_d$ están incorrelacionadas y esto implica que la matriz de varianzas del vector $\mathbf{x} = (x_1, \ldots, x_d)$ es diagonal.
   - El recíproco no es siempre cierto en general, pero si $x$ es gaussiano se cumple, pues

$$f_{\mathbf{x}}(x) = \prod_{i=1}^{d} f_{x_i}(x_i)$$

# Normal multivariada (teoremas)

Se puede probar:

1. Si $\mathbf{x} = (x_1, \dots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \dots, x_d$ independientes si y sólo si son incorrelacionadas (es decir $\Sigma$ es una matriz diagonal).

    - Es bien conocido que si $x_1, \dots, x_d$ son independientes entonces $x_1, \dots, x_d$ están incorrelacionadas y esto implica que la matriz de varianzas del vector $\mathbf{x} = (x_1, \dots, x_d)$ es diagonal.
    - El recíproco no es siempre cierto en general, pero si $x$ es gaussiano se cumple, pues

$$f_{\mathbf{x}}(x) = \prod_{i=1}^{d} f_{x_i}(x_i)$$

2. Si $\mathbf{x}_1, \dots, \mathbf{x}_n$ es una sucesión de vectores aleatorios i.i.d, con $E(\mathbf{x}_1) = \mu$ y matriz de covarianzas $\Sigma$ entonces

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{x}_i - \mu) \xrightarrow[n]{D} N_d(0_{\mathbb{R}^d}, \Sigma)$$

# Distribución normal bivariada

La distribución normal de un vector $(x_1, x_2)$ de media $\mu = (\mu_1, \mu_2)$ y matriz de covarianza

$$\Sigma = \begin{pmatrix} \sigma_1^2 & cov(x_1, x_2) \\ cov(x_2, x_1) & \sigma_2^2 \end{pmatrix}$$

tiene como densidad

$$f_{(x_1,x_2)}(x_1, x_2) = \frac{1}{(\sqrt{2\pi})^2 \sqrt{\det(\Sigma)}} \exp\left( -\frac{1}{2}(x_1 - \mu_1 \;\; x_2 - \mu_2)\Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right)$$

# Normal multivariada

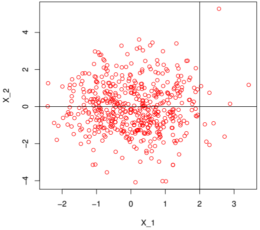



Figura: $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$\mathbb{P}(X_2 > 0 | X_1 > 2) = \mathbb{P}(X_2 > 0)$

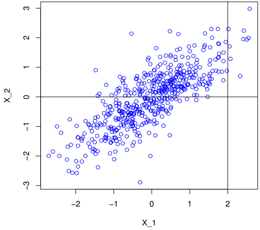



Figura: $\Sigma = \begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix}$

$\mathbb{P}(X_2 > 0 | X_1 > 2) > \mathbb{P}(X_2 > 0)$

# Normal Multivariada (propiedades)

Sea $\mathbf{x} \sim N(\mu, \Sigma)$ entonces tenemos las siguientes propiedades:

1. La distribución es simétrica alrededor de $\mu$.
   Esto es porque $f(\mu + a) = f(\mu - a)$.
2. La distribución tiene un único máximo en $\mu$.
   Al ser $\Sigma$ definida positiva, el término del exponente $(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$ es siempre positivo, y $f(\mathbf{x})$ es máxima cuando este término es nulo, o sea si $\mathbf{x} = \mu$.
3. La media es $\mu$ y la matriz de varianzas-covarianzas es $\Sigma$.
4. Las distribuciones marginales son normales.
5. Si $\mathbf{x} \in \mathbb{R}^d$ es normal y $A \in \mathcal{M}_{k \times d}$ es una matriz entonces $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^k$ es normal.

# Test normalidad en R

Nos preguntamos si los datos tienen distribución normal. Para eso podemos recurrir a la adaptación del test de normalidad de Shapiro-Wilk para datos multivariados.

```
library(mvnormtest)
x=iris[,1:4]
x=as.matrix(x)
mshapiro.test(t(x))
```

# Normal Multivariada

Si cortamos con hiperplanos paralelos a las $p$ variables, se obtienen curvas de nivel [2] cuya ecuación es

$$(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = cte \quad \text{(elipsoides)}$$

y define entonces una medida de la distancia de **x** al centro $\mu$.
Esta medida se llama distancia de Mahalanobis y

$$D^2 = (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$$

Supongamos que tenemos dos normales $N(0,1)$ y $N(10,10^2)$. Sea $x=3$, entonces:

1. La distancia euclidea de $x=3$ a 0 es más corta que la distancia de 3 a 10.
2. La distancia de Mahalanobis de $x=3$ a la distribución que tiene desviación típica 1 es $(3-0)1(3-0)=9$
3. La distancia de Mahalanobis de $x=3$ a la distribución que tiene desviación típica 10 es $(3-10)\frac{1}{100}(3-10)=0{,}49$

Con la distancia de Mahalanobis, el punto $x=3$ está más cerca de la segunda distribución. Es más probable que provenga de ella (lo veremos en breve).

[2] La curva de nivel $\alpha$ de una funcón $f:\mathbb{R}^d\to\mathbb{R}$ se define como $\mathcal{C}_\alpha = \{\mathbf{x}\in\mathbb{R}^d : f(\mathbf{x})=\alpha\}$

## Plan

# Distribuciones mezcladas

Muchas veces los datos multivariantes provienen de poblaciones distintas. Si suponemos que tenemos $G$ poblaciones, entonces

$$\mathbb{P}(\mathbf{x} \in A) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A, \mathbf{x} \in i) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A | \mathbf{x} \in i)\mathbb{P}(\mathbf{x} \in i) = \sum_{i=1}^{G} \pi_i \mathbb{P}(\mathbf{x} \in A | \mathbf{x} \in i)$$

y por lo tanto la función de densidad de la población es

$$f(\mathbf{x}) = \sum_{i=1}^{G} \pi_i f_i(\mathbf{x})$$

# Distribuciones mezcladas

Muchas veces los datos multivariantes provienen de poblaciones distintas. Si suponemos que tenemos $G$ poblaciones, entonces

$$\mathbb{P}(\mathbf{x} \in A) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A, \mathbf{x} \in i) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A | \mathbf{x} \in i)\mathbb{P}(\mathbf{x} \in i) = \sum_{i=1}^{G} \pi_i \mathbb{P}(\mathbf{x} \in A | \mathbf{x} \in i)$$

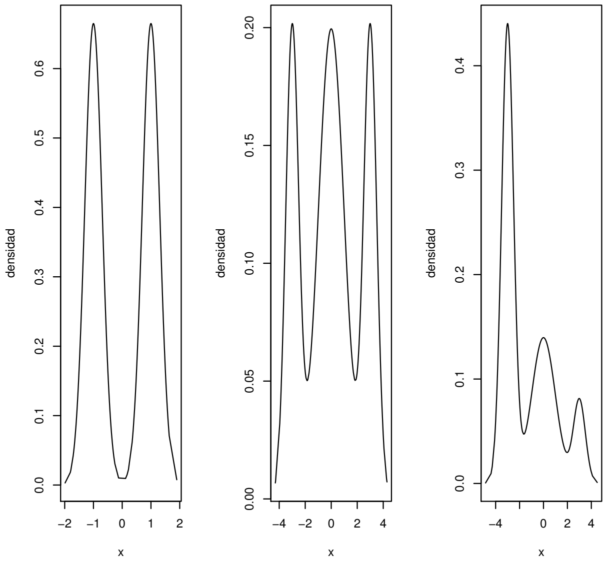
y por lo tanto la función de densidad de la población es

$$f(\mathbf{x}) = \sum_{i=1}^{G} \pi_i f_i(\mathbf{x})$$

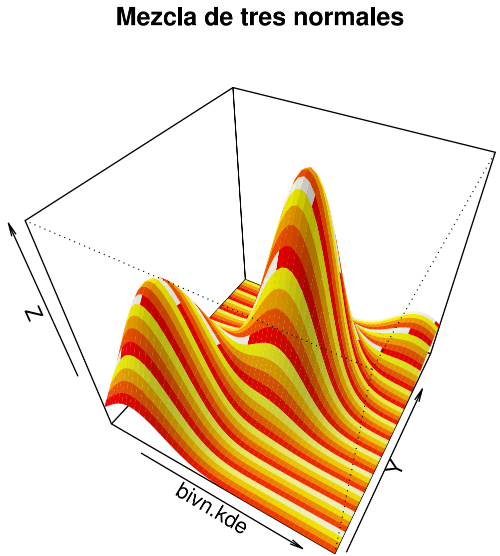
Se prueba que

1. La media de una mezcla es
$$\mu = \sum_{i=1}^{G} \pi_i \mu_i$$
.
2. La matriz de varianzas y covarianzas de una mezcla es
$$V = \sum_{i=1}^{G} \pi_i V_i + \sum_{i=1}^{G} \pi_i (\mu_i - \mu)(\mu_i - \mu)'$$
donde $V_i$ es la matriz de varianzas y covarianzas de la distribución $i$.

# Mezcla de normales multivariadas



# Mezcla de normales multivariadas



# Referencias


- D. Peña, *Analisis de Datos Multivariantes*, Mac Graw Hill, 2002.
- A. I. Izenman, *Modern Multivariate Statistical Techniques*, Springer, 2008.