Tema 1: Distribución normal multivariante

Amparo Baíllo Moreno

Departamento de Matemáticas
Universidad Autónoma de Madrid

Vectores aleatorios

Los datos multivariados son el resultado de observar un *vector aleatorio*, es decir, un vector $\mathbf{X} = (X_1, \ldots, X_p)'$ cuyas componentes X_j, $j = 1, \ldots, p$, son variables aleatorias (v.a.) definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$:

$$\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p}) \quad \text{medible.}$$

Una matriz aleatoria es una matriz cuyas componentes son v.a.

La distribución de probabilidad $\mathrm{P}_{\mathbf{X}}$ de un vector aleatorio $\mathbf{X}$ es la medida de probabilidad que $\mathbf{X}$ induce en $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$ y está caracterizada por la *función de distribución* F de $\mathbf{X}$:

$$F(\mathbf{x}) = \mathbb{P}\{X_1 \leq x_1, \ldots, X_p \leq x_p\}, \quad \text{para } \mathbf{x} = (x_1, \ldots, x_p)' \in \mathbb{R}^p.$$

La *función característica* de $\mathbf{X}$ es $\varphi(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}'\mathbf{X}})$, donde $\mathbf{t} \in \mathbb{R}^p$. Esta función caracteriza la distribución de $\mathbf{X}$.

La distribución de probabilidad $\mathrm{P}_{\mathbf{X}}$ de un vector aleatorio $\mathbf{X}$ continuo se caracteriza también mediante su *función de densidad*:

$$f(\mathbf{x}) = \frac{\partial^p F}{\partial x_1 \cdots \partial x_p} \quad \text{para } \mathbf{x} = (x_1, \ldots, x_p)' \in \mathbb{R}^p,$$

en el sentido de que

$$\mathbb{P}\{\mathbf{X} \in B\} = \mathrm{P}_{\mathbf{X}}(B) = \int_B f(\mathbf{x})\, d\mathbf{x}, \quad \text{para } B \in \mathcal{B}_{\mathbb{R}^d}.$$

Entonces se cumple que

$$\varphi(\mathbf{t}) = \int_{\mathbb{R}^p} e^{i\mathbf{t}'\mathbf{x}} f(\mathbf{x})\, d\mathbf{x}.$$

La *esperanza* o *media poblacional* de un vector aleatorio (resp. matriz) es el vector (resp. matriz) de esperanzas de cada una de sus componentes (las esperanzas marginales):

$$\boldsymbol{\mu} := \mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_p))' = (\mu_1, \ldots, \mu_p)'.$$

La esperanza es un operador lineal:

1. Si $\mathbf{A}$, matriz $q \times p$, y $\mathbf{b} \in \mathbb{R}^q$, son ambos deterministas y $\mathbf{X}$ es un vector aleatorio p-dimensional, entonces $\mathbb{E}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\mathbb{E}(\mathbf{X}) + \mathbf{b}$.
2. Si $\mathbf{X}$ e $\mathbf{Y}$ son matrices aleatorias de la misma dimensión, entonces $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$.
3. Si $\mathbf{X}$ es una matriz aleatoria $q \times p$ y $\mathbf{A}$, $\mathbf{B}$ son matrices deterministas de dimensiones adecuadas, entonces $\mathbb{E}(\mathbf{AXB}) = \mathbf{A}\mathbb{E}(\mathbf{X})\mathbf{B}$.

Si $\mathbf{X}_1$ y $\mathbf{X}_2$ son matrices aleatorias conformables e independientes, entonces $\mathbb{E}(\mathbf{X}_1\mathbf{X}_2) = \mathbb{E}(\mathbf{X}_1)\mathbb{E}(\mathbf{X}_2)$.

La *matriz de (varianzas-)covarianzas* de $\mathbf{X}$ es

$$\begin{aligned}\boldsymbol{\Sigma} = \mathbb{V}(\mathbf{X}) &:= \mathbb{E}((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})') = \mathbb{E}(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}' \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix},\end{aligned}$$

con $\sigma_{jj} = \mathbb{V}(X_j)$ y $\sigma_{jk} = \text{Cov}(X_j, X_k)$, para $j, k = 1, \dots, p$.

Propiedades de la matriz de covarianzas:

1. $\boldsymbol{\Sigma}$ es una matriz simétrica.
2. Si $\mathbf{A}$, matriz $q \times p$, y $\mathbf{b} \in \mathbb{R}^q$, son deterministas, entonces

$$\mathbb{V}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathbb{V}(\mathbf{X})\mathbf{A}'.$$

3. $\boldsymbol{\Sigma} = \mathbb{V}(\mathbf{X})$ es siempre definida no negativa.

Sea $\mathbf{D} := \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$. Entonces la *matriz de correlaciones* de $\mathbf{X}$ es

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1p} \\ \rho_{21} & 1 & \ldots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \ldots & 1 \end{pmatrix} = \mathbf{D}^{-1/2} \boldsymbol{\Sigma} \mathbf{D}^{-1/2},$$

donde ρ_{jk} es la correlación de X_j y X_k, para $j, k = 1, \ldots, p$, y

$$\mathbf{D}^{-1/2} := \mathrm{diag}(\sigma_{11}^{-1/2}, \ldots, \sigma_{pp}^{-1/2}).$$

Comprobar que, si $\mathbf{Z} := \mathbf{D}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, entonces $\mathbb{V}(\mathbf{Z}) = \boldsymbol{\rho}$.

Estandarización multivariante

Una matriz definida no negativa (como $\mathbf{\Sigma}$) tiene una única raíz cuadrada ($\mathbf{\Sigma}^{1/2}$) que sea definida no negativa.

Descomposición espectral de la matriz simétrica $\mathbf{\Sigma}$:
$\mathbf{\Sigma} = \mathbf{CDC}'$, con $\mathbf{C}$ matriz ortogonal cuyas columnas son los autovectores de $\mathbf{\Sigma}$ ortonormalizados y $\mathbf{D}$ matriz diagonal con los autovalores de $\mathbf{\Sigma}$ en la diagonal.

Entonces $\mathbf{\Sigma}^{1/2} := \mathbf{CD}^{1/2}\mathbf{C}'$ y $\mathbf{\Sigma}^{-1/2} := \mathbf{CD}^{-1/2}\mathbf{C}'$.

Dado $\mathbf{X}$ p-dimensional con $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ y $\mathbb{V}(\mathbf{X}) = \mathbf{\Sigma}$, definimos la *estandarización multivariante* de $\mathbf{X}$ como $\mathbf{Y} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$.

Geométricamente, la estandarización $\mathbf{Y} = \mathbf{CD}^{-1/2}\mathbf{C}'(\mathbf{X} - \boldsymbol{\mu})$ equivale a las siguientes operaciones sobre $\mathbf{X}$:

- Trasladarlos para que la media sea el origen (centrar $\mathbf{X}$).
- Rotarlos de forma que las correlaciones se anulen.
- Cambios de escala en cada variable, para que las varianzas sean 1.
- Deshacer la rotación anterior (sin efectos sobre la matriz de covarianzas).

La distribución normal multivariante

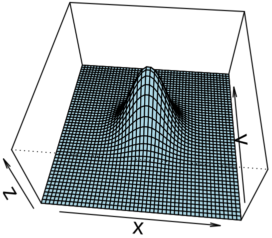
El vector aleatorio $\mathbf{X}$ es *normal* p-dimensional con parámetros $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma}$, matriz $p \times p$ definida positiva, si su densidad es:

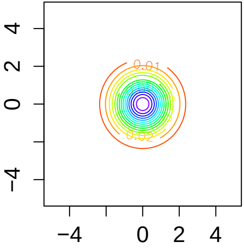
$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \quad \mathbf{x} \in \mathbb{R}^p.$$

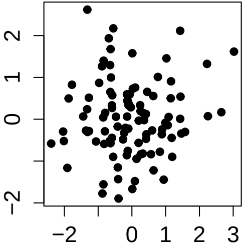
Lo denotaremos $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Ejemplos de densidades normales bidimensionales ($p = 2$):

$$\boldsymbol{\mu} = \mathbf{0} = (0,0)' \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \text{diag}(1,1) = \mathbf{I}$$



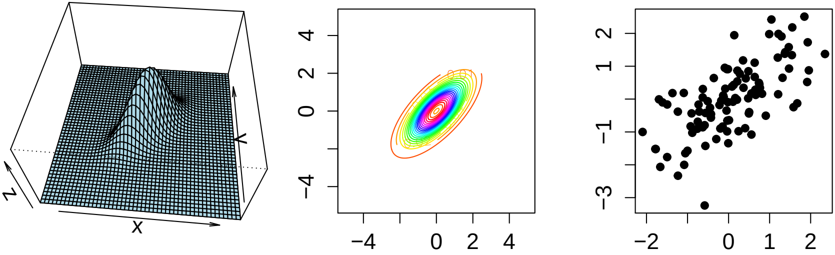




```
split.screen(c(1,3))

screen(1)
## bivariate normal pdf
library(mvtnorm)
x = y = seq(-5, 5, length = 50)
f = function(x,y) { dmvnorm(cbind(x,y)) }
z = outer(x, y, f)
par(mai=c(0.1,0.1,0.1,0.1))
persp(x, y, z, theta=5, phi=50, expand=0.5, col="lightblue")

screen(2)
## contours of the bivariate normal pdf
x = y = seq(-5, 5, length = 150)
z = outer(x, y, f)
par(mai=c(0.5,0.5,0.5,0.5))
contour(x, y, z, nlevels=20, col=rainbow(20))

screen(3)
## normal data
X = rmvnorm(n=100,sigma=matrix(c(1,0,0,1),ncol=2))
par(mai=c(0.5,0.5,0.5,0.5))
plot(X[,1],X[,2], pch=19,xlab=expression(x[1]),ylab=expression(x[2]))
```

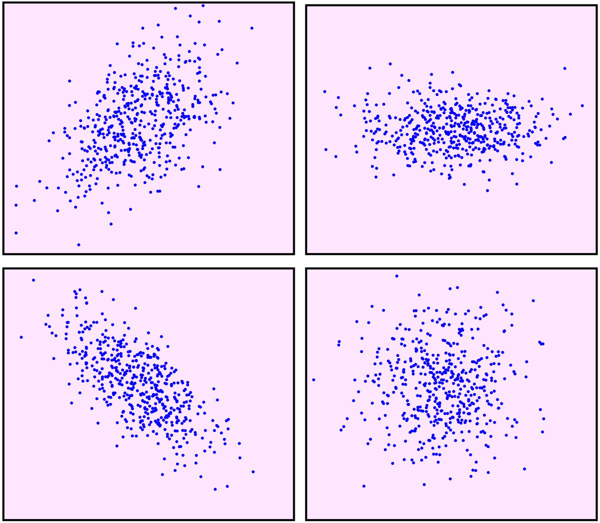
Ejemplos de densidades normales bidimensionales ($p = 2$):

$$\boldsymbol{\mu} = \mathbf{0} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$$



```
screen(1)
x = y = seq(-5, 5, length = 50)
Sigma = matrix(c(1,0.7,0.7,1), ncol=2)
f = function(x,y) { dmvnorm(cbind(x,y),sigma=Sigma) }
z = outer(x, y, f)
par(mai=c(0.1,0.1,0.1,0.1))
persp(x, y, z, theta=5, phi=50, expand=0.5, col="lightblue")

screen(2)
## contours of the bivariate normal pdf
x = y = seq(-5, 5, length = 150)
z = outer(x, y, f)
par(mai=c(0.5,0.5,0.5,0.5))
contour(x, y, z, nlevels=20, col=rainbow(20))

screen(3)
## normal data
X = rmvnorm(n=100,sigma=Sigma)
par(mai=c(0.5,0.5,0.5,0.5))
plot(X[,1],X[,2], pch=19,xlab=expression(x[1]),ylab=expression(x[2]))
```

Relaciona cada matriz con su conjunto de datos:



$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_4 = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$$

Proposición: Sea $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y sea $\mathbf{Y} = (Y_1, \ldots, Y_p)' = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ su estandarización multivariante. Entonces $Y_1, \ldots, Y_p$ son i.i.d. $\mathrm{N}(0, 1)$.

Dem: Cambio de variable $\mathbf{h}(\mathbf{x}) = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ y $|\mathbf{J}_{\mathbf{h}^{-1}}| = |\boldsymbol{\Sigma}|^{1/2}$.

Consecuencias:

1. Si $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ y $\mathbb{V}(\mathbf{X}) = \boldsymbol{\Sigma}$.

(puesto que $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Y} + \boldsymbol{\mu}$).

2. Si $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces $\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\{i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}$

(ya que $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Y} + \boldsymbol{\mu}$ y la f.c. de una v.a. normal estándar es $e^{-t^2/2}$).

3. La distribución de $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ es χ^2_p

(puesto que $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Y}'\mathbf{Y}$).

Distancia de Mahalanobis

Si $\mathbf{X} = (X_1, \ldots, X_p)'$ es un vector con media $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$, la *distancia de Mahalanobis* de $\mathbf{X}$ a $\boldsymbol{\mu}$ es

$$d_M(\mathbf{X}, \boldsymbol{\mu}) = \sqrt{(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})}.$$

En general, la distancia de Mahalanobis entre dos vectores $\mathbf{x}$ e $\mathbf{y}$ (respecto a una matriz de varianzas covarianzas $\boldsymbol{\Sigma}$) es

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{y})}.$$

Distancia de Mahalanobis frente a otras distancias:

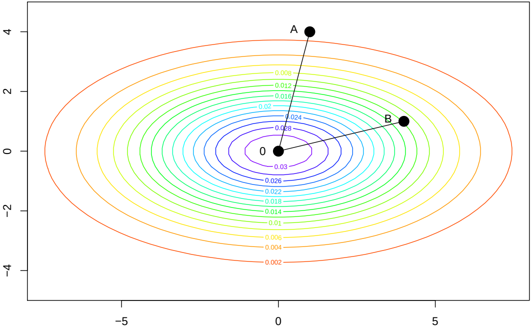
1. d_M coincide con la distancia euclídea d_E entre los datos estandarizados de forma multivariante.
2. d_M es adimensional.
3. d_M tiene en cuenta las diferentes variabilidades (varianzas) de las variables.
4. d_M tiene en cuenta las correlaciones entre las variables.
5. Bajo normalidad, su cuadrado se distribuye como una χ^2_p.

Ejemplo:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 10 & 0 \\ 0 & 2.5 \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.4 \end{bmatrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

- $d_M^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{y}) = 0.1(x_1 - y_1)^2 + 0.4(x_2 - y_2)^2$.
- $d_E(A, O) = d_E(B, O)$, $d_M(A, O) > d_M(B, O)$.
- d_M tiene en cuenta la diferente variabilidad de las componentes.

$A = (1, 4)$
$B = (4, 1)$

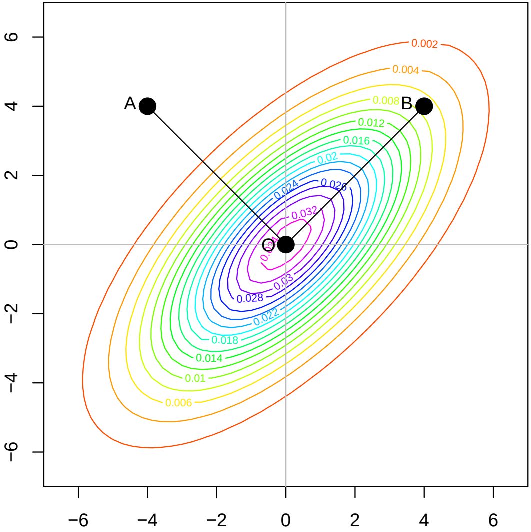


```
library(mvtnorm)
x = seq(-8, 8, length = 50)
y = seq(-5, 5, length = 50)
Sigma = matrix(c(10,0,0,2.5), ncol=2)
f = function(x,y) { dmvnorm(cbind(x,y),sigma=Sigma) }
z = outer(x, y, f)

pdf(file="MahalDistExample.pdf",width=8,height=5)

par(mai=c(0.5,0.5,0.1,0.1))
contour(x, y, z, nlevels=20, col=rainbow(20), xaxs="i",yaxs="i",xlim=c(-8,8),ylim=c(-5,5))
points(1,4,pch=19,cex=2)
segments(0,0,1,4)
points(4,1,pch=19,cex=2)
segments(0,0,4,1)
points(0,0,pch=19,cex=2)
text(0.5,4.1,"A")
text(3.5,1.1,"B")
text(-0.5,0,"O")

dev.off()

A = matrix(c(1,4),ncol=1)
B = matrix(c(4,1),ncol=1)
MDAO = t(A) %*% solve(Sigma) %*% A
MDBO = t(B) %*% solve(Sigma) %*% B
```

Ejemplo:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 0.3 & -0.2 \\ -0.2 & 0.3 \end{bmatrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}.$$

- $A = (-4, 4),\ B = (4, 4)$
- $d_E(A, O) = d_E(B, O)$
- $d_M(A, O) > d_M(B, O)$
- d_M incluye la correlación



```
x = seq(-7, 7, length = 50)
y = seq(-7, 7, length = 50)
Sigma = matrix(c(6,4,4,6), ncol=2)
f = function(x,y) { dmvnorm(cbind(x,y),sigma=Sigma) }
z = outer(x, y, f)

pdf(file="MahalDistExample2.pdf",width=8,height=5)

par(mai=c(0.5,0.5,0.1,0.1),pty="s")
contour(x, y, z, nlevels=20, col=rainbow(20), xaxs="i",yaxs="i")
abline(h=0,col="gray")
abline(v=0,col="gray")
points(-4,4,pch=19,cex=2)
segments(0,0,-4,4)
points(4,4,pch=19,cex=2)
segments(0,0,4,4)
points(0,0,pch=19,cex=2)
text(-4.5,4.1,"A")
text(3.5,4.1,"B")
text(-0.5,0,"O")

dev.off()

A = matrix(c(-4,4),ncol=1)
B = matrix(c(4,4),ncol=1)
MDAO = t(A) %*% solve(Sigma) %*% A
MDBO = t(B) %*% solve(Sigma) %*% B
```

Muestra y estimación

Dada una muestra $\mathbf{X}_1, \ldots, \mathbf{X}_n$ de $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, los estimadores de máxima verosimilitud (e.m.v.) de $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ son respectivamente

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} := \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$$

y

$$\hat{\boldsymbol{\Sigma}} = \mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' = \begin{bmatrix} S_{11} & S_{12} & \ldots & S_{1p} \\ S_{21} & S_{22} & \ldots & S_{2p} \\ \vdots & \vdots & & \vdots \\ S_{p1} & S_{p2} & \ldots & S_{pp} \end{bmatrix},$$

siendo $\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$ la media muestral de la j-ésima variable, X_j, y $S_{jk} = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$ la covarianza muestral entre X_j y X_k.

Otro estimador de $\boldsymbol{\Sigma}$ es $\mathbf{S}_{n-1} = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$.

Ejemplo: Depósitos de corcho (en cg.) para 28 alcornoques en las cuatro direcciones cardinales

N	E	S	W
72	66	76	77
60	53	66	63
56	57	64	58
41	29	36	38
32	32	35	36
30	35	34	26
39	39	31	27
42	43	31	25
37	40	31	25
33	29	27	36
32	30	34	28
63	45	74	63
54	46	60	52
47	51	52	43
91	79	100	75
56	68	47	50
79	65	70	61
81	80	68	58
78	55	67	60
46	38	37	38
39	35	34	37
32	30	30	32
60	50	67	54
35	37	48	39
39	36	39	31
50	34	37	40
43	37	39	50
48	54	57	43

```
Datos = read.table("alcornoques.txt",header=TRUE)
colMeans(Datos)
       N        E        S        W
50.53571 46.17857 49.67857 45.17857
```

$$\bar{\mathbf{x}} = \begin{bmatrix} 50.53571 \\ 46.17857 \\ 49.67857 \\ 45.17857 \end{bmatrix}$$

Tanto `var(Datos)` como

```
cov(Datos)
         N        E        S        W
N 290.4061 223.7526 288.4378 226.2712
E 223.7526 219.9299 229.0595 171.3743
S 288.4378 229.0595 350.0040 259.5410
W 226.2712 171.3743 259.5410 226.0040
```

calculan $\mathbf{S}_{n-1}$.

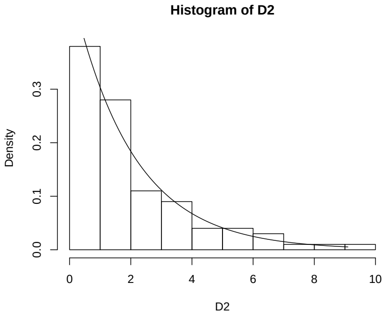
Entonces, para $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, podemos estimar $d_M(\mathbf{x}, \mathbf{y})$ (la distancia de Mahalanobis poblacional entre $\mathbf{x}$ y $\mathbf{y}$) mediante su análogo muestral

$$\hat{d}_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}$$

y $d_M(\mathbf{x}, \boldsymbol{\mu})$ mediante $\hat{d}_M(\mathbf{x}, \bar{\mathbf{x}}) = \sqrt{(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})}$.

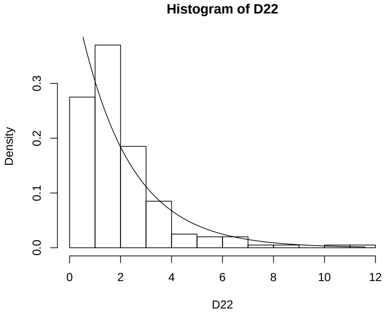
Si una muestra $\mathbf{x}_1, \ldots, \mathbf{x}_n$ procede de una distribución normal, entonces $d_M^2(\mathbf{x}_i, \boldsymbol{\mu})$, $i = 1, \ldots, n$, sigue una distribución χ_p^2. Por tanto, $\hat{d}_M^2(\mathbf{x}_i, \bar{\mathbf{x}})$ debería seguir aproximadamente una χ^2.

```
n = 100 # tamaño muestral
Sigma = matrix(c(1,0,0,1), ncol=2)
# Muestra de una normal bivariante estándar:
X = rmvnorm(n,mean=c(0,0),sigma=Sigma)
m = colMeans(X) # medias muestrales
S = cov(X)
# Cuadrados de distancias de Mahalanobis a m:
D2 = mahalanobis(X, m, S)

pdf(file="MahalDistChi2.pdf",width=6,height=5)
  hist(D2,freq=F)
  ejex = seq(0,max(D2),0.1)
  denschi2 = dchisq(ejex,df=2)
  lines(ejex,denschi2)
dev.off()
```



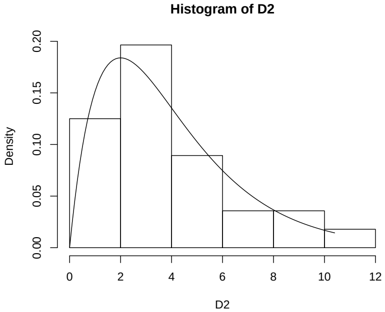
Ejemplo con una muestra no normal:

```
X2 = rbind(X,rmvnorm(n,mean=c(3,3),sigma=Sigma) )
m2 = colMeans(X2)
S2 = cov(X2)
D22 = mahalanobis(X2, m2, S2)

pdf(file="MahalDistChi22.pdf",width=6,height=5)
hist(D22,freq=F)
ejex = seq(0,max(D22),0.1)
denschi2 = dchisq(ejex,df=2)
lines(ejex,denschi2)
dev.off()
```



Ejemplo datos alcornoques:

```
Datos = read.table("alcornoques.txt",header=TRUE)
m = colMeans(Datos)
S = cov(Datos)
D2 = mahalanobis(Datos, m, S)

pdf(file="MahalDistAlcor.pdf",width=6,height=5)
hist(D2,freq=F)
ejex = seq(0,max(D2),0.1)
denschi2 = dchisq(ejex,df=4)
lines(ejex,denschi2)
dev.off()
```



Generalmente uno de los objetivos de la Estadística Multivariante es entender la estructura de dependencias entre las variables observadas, $X_1, \ldots, X_p$. Estas dependencias pueden ser

- entre pares de variables;
- entre una variable y el resto;
- global (entre todas las variables); ...

La *dependencia lineal* entre parejas de variables se describe mediante la *matriz de correlaciones muestrales*

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \ldots & r_{1p} \\ r_{21} & 1 & \ldots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \ldots & 1 \end{bmatrix},$$

donde $r_{jk} = \dfrac{s_{jk}}{\sqrt{s_{jj}s_{kk}}}$ es el *coeficiente de correlación muestral* entre X_j y X_k.

Las matrices de correlación y de covarianzas están relacionadas de la siguiente manera

$$\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2},$$

donde $\mathbf{D}$ es la matriz diagonal cuya diagonal principal es igual a la de $\mathbf{S}$:

$$\mathbf{D} = \begin{bmatrix} s_{11} & 0 & \dots & 0 \\ 0 & s_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & s_{pp} \end{bmatrix} \quad \text{y} \quad \mathbf{D}^{-1/2} = \begin{bmatrix} s_{11}^{-1/2} & 0 & \dots & 0 \\ 0 & s_{22}^{-1/2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & s_{pp}^{-1/2} \end{bmatrix}$$

R es la matriz de covarianzas muestrales de la muestra estandarizada

$$\frac{x_{i1} - \bar{x}_1}{\sqrt{s_{11}}}, \frac{x_{i2} - \bar{x}_2}{\sqrt{s_{22}}}, \ldots, \frac{x_{ip} - \bar{x}_p}{\sqrt{s_{pp}}}, \quad i = 1, \ldots, n.$$

Ejemplo datos alcornoques: Con R

```
cor(Datos)
          N         E         S         W
N 1.0000000 0.8853667 0.9047173 0.8832188
E 0.8853667 1.0000000 0.8256001 0.7686801
S 0.9047173 0.8256001 1.0000000 0.9228082
W 0.8832188 0.7686801 0.9228082 1.0000000
```

Transformaciones afines de vectores normales

Proposición: Si $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{A}$ es matriz constante $q \times p$ y $\mathbf{b} \in \mathbb{R}^q$ es un vector constante, entonces

$$\mathbf{AX} + \mathbf{b} \sim \mathrm{N}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

Dem: Ejercicio (usar la f.c.)

Consecuencia: Si $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$, con $\mathbf{X}_1 \in \mathbb{R}^q$ y $\mathbf{X}_2 \in \mathbb{R}^{p-q}$, y consideramos las particiones correspondientes de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$,

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right)$$

entonces $\mathbf{X}_1 \sim \mathrm{N}_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$. En general, todos los subconjuntos de componentes de $\mathbf{X}$ siguen una distribución normal.

Proposición: Sea $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$. Entonces $\mathbf{X}_1$ y $\mathbf{X}_2$ son independientes $\Longleftrightarrow \boldsymbol{\Sigma}_{12} = 0$.

Dem: Ejercicio (descomponer la densidad de $\mathbf{X}$ como producto de las densidades de $\mathbf{X}_1$ y $\mathbf{X}_2$).

Observaciones:

- Si dos v.a. X e Y tienen distribución normal y además $\mathrm{Cov}(X, Y) = 0$, esto no implica que X e Y sean independientes.
- Si dos v.a. X e Y tienen distribución normal y $a, b \in \mathbb{R}$, la combinación lineal $aX + bY$ no tiene necesariamente distribución normal.
- Aunque todas las marginales de un vector aleatorio p-dimensional $\mathbf{X}$ tengan distribución normal, esto no implica que $\mathbf{X}$ tenga distribución normal p-dimensional.

Ejercicio: Añade las hipótesis necesarias para que las conclusiones sí sean válidas.

Ejemplo: $\mathbf{X} = (X_1, X_2, X_3)' \sim \mathrm{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 7/2 & 1/2 & -1 \\ 1/2 & 1/2 & 0 \\ -1 & 0 & 1/2 \end{pmatrix}.$$

a) Calcula las distribuciones marginales.

b) Calcula la distribución del vector $(X_1, X_2)'$.

c) ¿Son X_2 y X_3 independientes?

d) ¿Es X_3 independiente del vector $(X_1, X_2)'$?

e) Calcula la distribución de la variable $2X_1 - X_2 + 3X_3$.

Distribuciones condicionadas

Proposición: Sea $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con $\mathbf{X}_1 \in \mathbb{R}^q$ y $\mathbf{X}_2 \in \mathbb{R}^{p-q}$. Consideramos las particiones correspondientes de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ y suponemos que $|\boldsymbol{\Sigma}_{11}| \neq 0$. Entonces,

$$\mathbf{X}_2 | \mathbf{X}_1 \sim \mathrm{N}_{p-q}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{2.1}),$$

donde

$$\begin{aligned} \boldsymbol{\mu}_{2.1} &= \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1), \\ \boldsymbol{\Sigma}_{2.1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}. \end{aligned}$$

Observaciones:

- $\boldsymbol{\mu}_{2.1} = \mathbb{E}(\mathbf{X}_2 | \mathbf{X}_1)$ es una función lineal (afín) de $\mathbf{X}_1$.
- $\boldsymbol{\Sigma}_{2.1}$ no depende de $\mathbf{X}_1$ (homocedasticidad).

Ejemplos:

Sea $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix} \right)$

1. Calcula la distribución de Y dada X.
2. Repite el ejercicio para la distribución de X dada Y.

Sea $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \right)$.

1. Calcula la distribución de $X + Y$ condicionada a que $X - Y = 1$.

Formas cuadráticas bajo normalidad

Proposición: Sea $\mathbf{Y} \sim \mathrm{N}_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ y sea $\mathbf{B}$ una matriz $n \times n$ simétrica e idempotente. Supongamos que $\boldsymbol{\mu}' \mathbf{B} \boldsymbol{\mu} = 0$. Entonces

$$\frac{1}{\sigma^2} \mathbf{Y}' \mathbf{B} \mathbf{Y} \sim \chi^2_p,$$

donde p es la traza de $\mathbf{B}$.

Proposición: Sea $\mathbf{Y} \sim \mathrm{N}_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ y sean $\mathbf{A}_{p \times n}$, $\mathbf{B}_{q \times n}$, $\mathbf{C}_{n \times n}$ y $\mathbf{D}_{n \times n}$ con $\mathbf{C}$ y $\mathbf{D}$ simétricas e idempotentes.

1. $\mathbf{AY}$ y $\mathbf{BY}$ son independientes $\Longleftrightarrow \mathbf{AB}' = \mathbf{0}$.
2. $\mathbf{Y}'\mathbf{CY}$ e $\mathbf{Y}'\mathbf{DY}$ son independientes $\Longleftrightarrow \mathbf{CD} = \mathbf{0}$.
3. $\mathbf{AY}$ e $\mathbf{Y}'\mathbf{CY}$ son independientes $\Longleftrightarrow \mathbf{AC} = \mathbf{0}$.

Algunas aplicaciones:

1. Si proyectamos $\mathbf{Y}$ en dos direcciones ortogonales, tanto las proyecciones como sus normas son independientes.

2. **Lema de Fisher**: Se demuestra a partir de la obs. siguiente.

   Sea $\mathbf{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$ y $\mathbf{M} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$ (una matriz $n \times n$ con todas sus entradas iguales a $1/n$), entonces

$$\bar{Y}\mathbf{1}_n = \mathbf{MY}$$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \mathbf{Y}'(\mathbf{I}_n - \mathbf{M})\mathbf{Y},$$

   donde $\bar{Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i$.
   Interpretación geométrica: $\mathbf{M}$ es una matriz de proyección sobre el subespacio de $\mathbb{R}^n$ cuya base es $\mathbf{1}_n$.

Teorema central del límite multivariante

TCL para variables aleatorias unidimensionales:

Sean $X_1, X_2, \dots$ v.a.i.i.d. con esperanza μ y varianza σ^2, entonces $\sqrt{n}(\bar{X}_n - \mu) \to_d \mathrm{N}(0, \sigma^2)$, donde $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Teorema central del límite para vectores aleatorios:

Sean $\mathbf{X}_1, \mathbf{X}_2, \dots$ vectores aleatorios p dimensionales i.i.d. con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$, entonces

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \to_d \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}), \text{ donde } \bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i.$$

Referencias

Izenman (2008). *Modern Multivariate Statistical Techniques*. Springer.
Cap. 3: "Random Vectors and Matrices"

Johnson, R.A., Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. Pearson-Prentice Hall.
Cap. 1: "Aspects of multivariate analysis"
Cap. 2: "Matrix algebra and random vectors"
Cap. 3: "Sample geometry and random sampling"
Cap. 4: "The multivariate normal distribution"