Estadística II
Tema 1: Distribución normal multivariante

José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

Algunas propiedades de los vectores aleatorios

Sea $X = (X_1, \ldots, X_p)'$ un vector aleatorio p-dimensional.

- Su **vector de medias** es $\mathbb{E}(X) = \mu = (\mu_1, \ldots, \mu_p)'$ donde $\mu_i = \mathbb{E}(X_i)$.
- Su **matriz de covarianzas** es $\mathrm{Var}(X) = \Sigma$, cuya posición (i, j) es $\sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$. Es fácil comprobar

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mu)(X - \mu)'] = \mathbb{E}(XX') - \mu\mu'.$$

Transformaciones afines: si A es matriz $q \times p$ y $b \in \mathbb{R}^q$,

- $\mathbb{E}(AX + b) = A\mu + b$.
- $\mathrm{Var}(AX + b) = \mathbb{E}[A(X - \mu)(X - \mu)'A'] = A\Sigma A'$.

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

La distribución normal multivariante

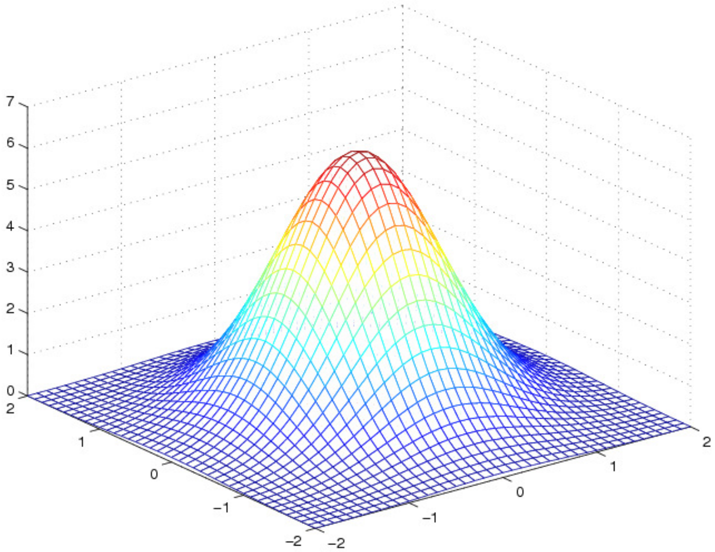
El vector aleatorio X es **normal** p-dimensional con vector de medias μ y matriz de covarianzas Σ (notación: $X \equiv \mathsf{N}_p(\mu, \Sigma)$) si tiene densidad dada por:

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

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}.$$

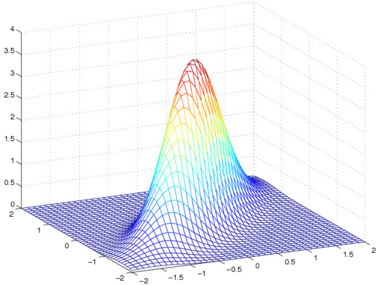
Ejemplos de densidades normales bidimensionales

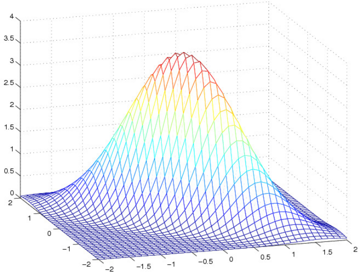
$$\mu = (0,0)' \text{ y } \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$



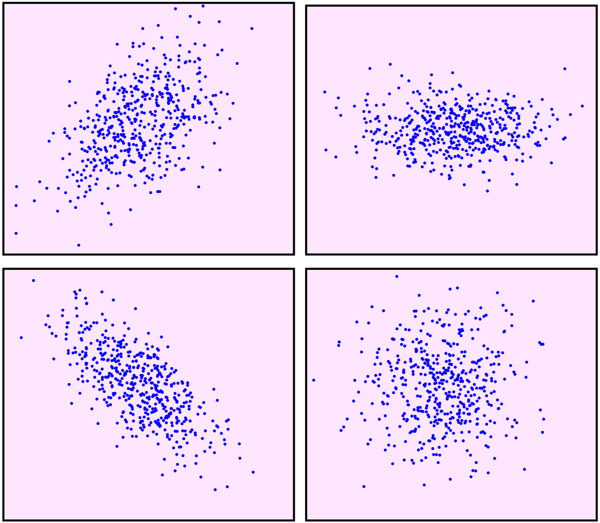
Ejemplos de densidades normales bidimensionales

$\mu = (0,0)'$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$ $\qquad$ $\mu = (0,0)'$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}$





Relaciona cada matriz con su conjunto de datos



$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}$$

$$\Sigma_3 = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}$$

$$\Sigma_4 = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$$

Estandarización multivariante

Prop: Si $X \equiv \mathrm{N}_p(\mu, \Sigma)$ y definimos $Y = \Sigma^{-1/2}(X - \mu)$, entonces $Y_1, \ldots, Y_p$ son i.i.d. $\mathrm{N}(0, 1)$.

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

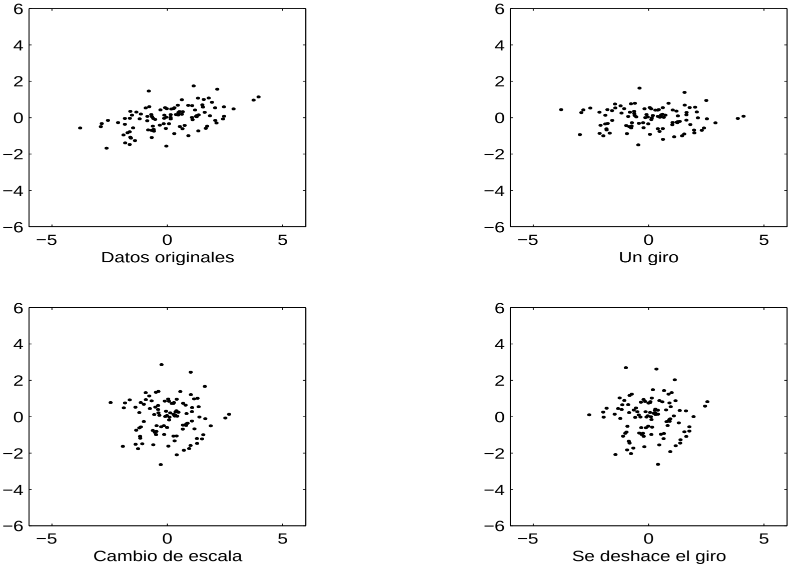
Σ simétrica $\Rightarrow \Sigma = CDC' \Rightarrow \Sigma^{-1/2} = CD^{-1/2}C'$. Por lo tanto,

$$Y = CD^{-1/2}C'(X - \mu).$$

Geométricamente, la estandarización equivale a las siguientes operaciones sobre los datos:

- Trasladarlos para que la media sea el origen.
- Rotarlos de forma que las correlaciones se anulen.
- Cambios de escala en cada variable, para que las varianzas sean 1.
- Deshacer la rotación anterior (no cambia la distribución).

Estandarización multivariante



Estandarización multivariante: consecuencias

1. Si $X \equiv \mathrm{N}_p(\mu, \Sigma)$, entonces $\mathbb{E}(X) = \mu$ y $\mathrm{Var}(X) = \Sigma$.

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

2. Si $X \equiv \mathrm{N}_p(\mu, \Sigma)$, entonces $\varphi_X(t) = \exp\{it'\mu - \frac{1}{2} t'\Sigma t\}$

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

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

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

Distancia de Mahalanobis

Si $X = (X_1, \ldots, X_p)'$ es un vector con media μ y covarianzas Σ, la **distancia de Mahalanobis** de X a μ es

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

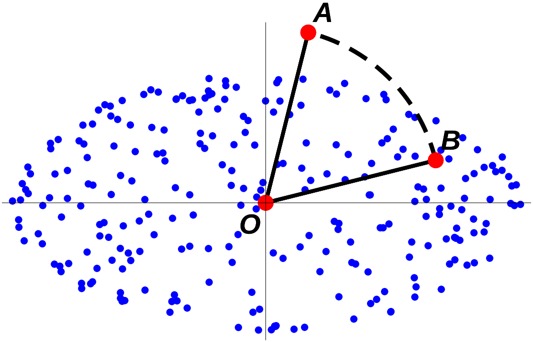
Distancia de Mahalanobis frente a otras distancias

1. d_M coincide con la distancia euclídea 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

$$\Sigma = \begin{bmatrix} 10 & 0 \\ 0 & 2,5 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0,1 & 0 \\ 0 & 0,4 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

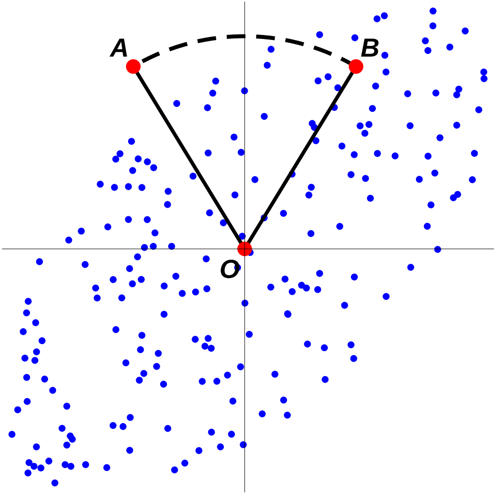
- $d_M^2(x,y) = (x-y)'\Sigma^{-1}(x-y) = 0,1(x_1-y_1)^2 + 0.4(x_2-y_2)^2.$
- $d_E(A,O) = d_E(B,O)$, $d_M(A,O) > d_M(B,O)$.
- d_M tiene en cuenta la diferente variabilidad de las variables.



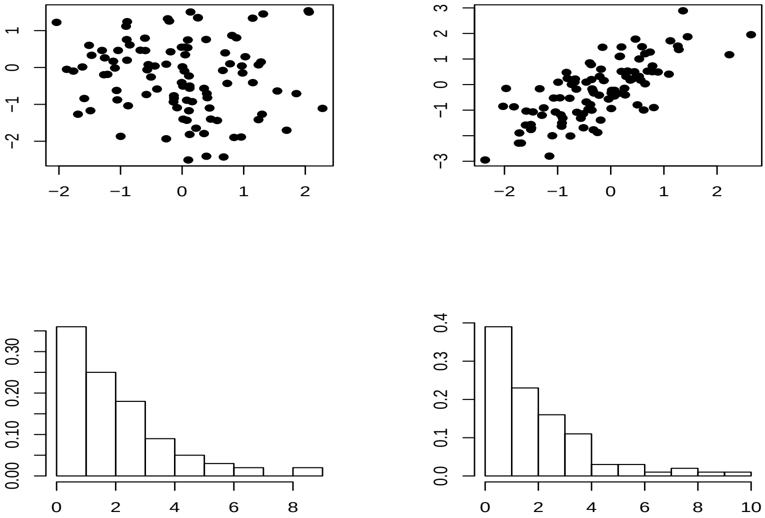
Ejemplo

$$\Sigma = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0,3 & -0,2 \\ -0,2 & 0,3 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0,6 \\ 0,6 & 1 \end{bmatrix}.$$

- $d_M^2 = (x-\mu)'\Sigma^{-1}(x-\mu)$.
- $d_E(A,O) = d_E(B,O)$.
- $d_M(A,O) > d_M(B,O)$.
- d_M incluye la correlación.



Distancias de Mahalanobis para datos normales



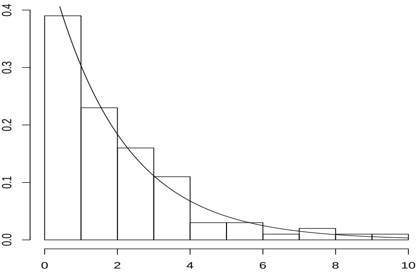
Distancias de Mahalanobis para datos normales

Estadísticos descriptivos para D_i^2 en el segundo ejemplo:

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.007255 0.565100 1.314000 1.980000 2.710000 9.735000

Desviacion tipica: 1.920563
```

Comparación con la densidad χ^2:



Datos de lirios de Fisher

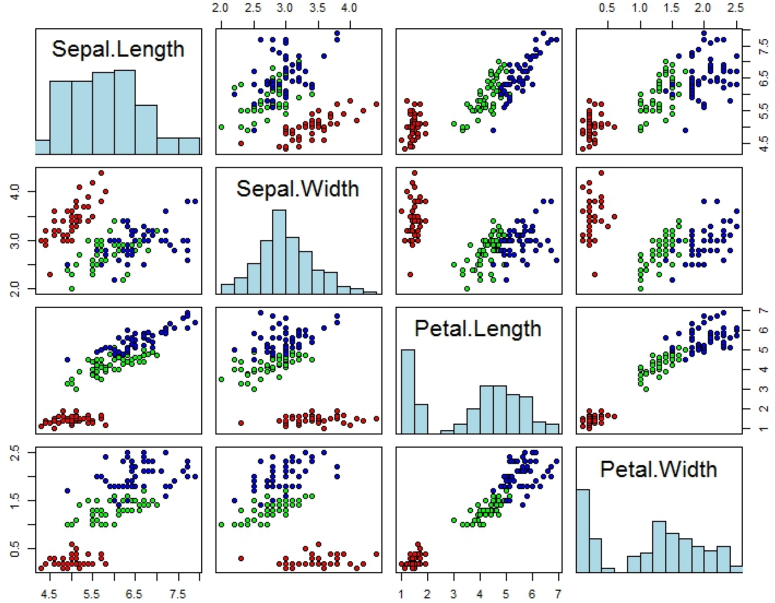
Conjunto de datos (iris data) muy famoso utilizado por Fisher (1936) para ilustrar un ejemplo de clasificación.

Enlace a la entrada en wikipedia sobre estos datos

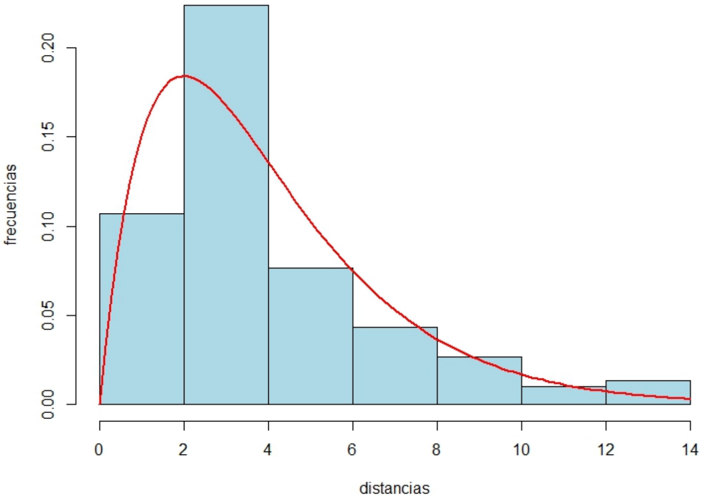
Medidas de la longitud y anchura del pétalo y el sépalo de 150 lirios

Hay 50 lirios de cada una de las especies *iris setosa*, *iris versicolor* e *iris virginica*.

Datos de lirios de Fisher



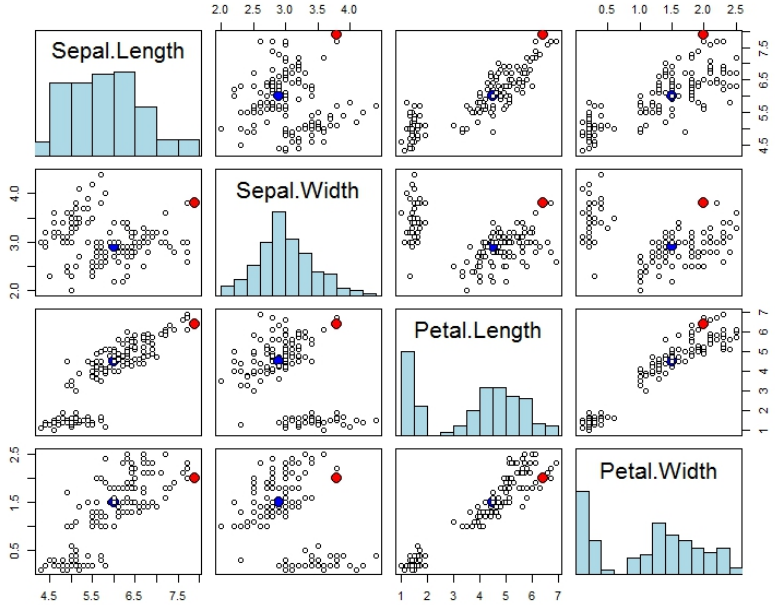
Distancias de Mahalanobis



Distancias de Mahalanobis

- La forma del histograma no coincide del todo con lo que se espera bajo normalidad (distribución χ^2)
- La distancia de Mahalanobis media es 3.97 y la varianza de las distancias es 7.69
- La mayor distancia es 13.1 y corresponde al número 132.
- La menor distancia es 0.32 y corresponde al número 79.

Datos de lirios de Fisher



Transformaciones afines de vectores normales

Prop: Si $X \equiv \mathsf{N}_p(\mu, \Sigma)$, A es matriz $q \times p$ y $b \in \mathbb{R}^q$, entonces

$$AX + b \equiv \mathsf{N}_q(A\mu + b, A\Sigma A')$$

Dem: Ejercicio (usar la f.c.)

Consecuencia: si $X = (X_1 \mid X_2)$, con $X_1 \in \mathbb{R}^q$ y $X_2 \in \mathbb{R}^{p-q}$, y consideramos las particiones correspondientes de μ y Σ,

$$\mu = (\mu_1 \mid \mu_2), \quad \Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right)$$

entonces $X_1 \equiv \mathsf{N}_q(\mu_1, \Sigma_{11})$.

Prop: X_1 y X_2 son independientes $\iff \Sigma_{12} = 0$.
Dem: Ejercicio (descomponer la densidad de X como producto de las densidades de X_1 y X_2).

Observaciones

- Si dos v.a. X e Y tienen distribución normal y además $\text{Cov}(X, Y) = 0$, 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 X tengan distribución normal, esto no implica que X tenga distribución normal p-dimensional.

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

Ejemplo

$X = (X_1, X_2, X_3)' \equiv \mathrm{N}_3(\mu, \Sigma)$ con

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

1. Calcula las distribuciones marginales.
2. Calcula la distribución del vector $(X_1, X_2)'$.
3. ¿Son X_2 y X_3 independientes?
4. ¿Es X_3 independiente del vector $(X_1, X_2)'$?
5. Calcula la distribución de la variable $2X_1 - X_2 + 3X_3$.

Distribuciones condicionadas

Prop: Sea $X = (X_1 \mid X_2)$, con $X_1 \in \mathbb{R}^q$ y $X_2 \in \mathbb{R}^{p-q}$. Consideramos las particiones correspondientes de μ y Σ y suponemos que existe Σ_{11}^{-1}. Entonces,

$$X_2|X_1 \equiv \mathrm{N}_{p-q}(\mu_{2.1}, \Sigma_{2.1}),$$

donde

$$\begin{aligned}
\mu_{2.1} &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \\
\Sigma_{2.1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.
\end{aligned}$$

- $\mu_{2.1} = \mathbb{E}(X_2|X_1)$ es una función lineal (afín) de X_1.
- $\Sigma_{2.1}$ no depende de X_1 (homocedasticidad).

Ejemplos

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

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

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

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

Formas cuadráticas bajo normalidad

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

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

donde p es la traza de B.

Prop: Sea $Y \equiv \mathrm{N}_n(\mu, \sigma^2 I_n)$ y sean $A_{p \times n}$, $B_{q \times n}$, $C_{n \times n}$ y $D_{n \times n}$ con C y D simétricas e idempotentes.

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

Algunas aplicaciones

1. Si proyectamos Y en dos direccciones ortogonales, tanto las proyecciones como sus normas son independientes.

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

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

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

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = Y'(I_n - M)Y.$$

   Interpretación geométrica: M es una matriz de proyección sobre el subespacio de $\mathbb{R}^n$ cuya base es 1_n.

Teorema central del límite multivariante

TCL para variables aleatorias unidimensionales:

Sean $X_1, X_2, \ldots$ v.a.i.i.d. con esperanza μ y varianza σ^2, entonces $\sqrt{n}(\bar{X}_n - \mu) \rightarrow_d \mathrm{N}(0, \sigma^2)$, donde $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$.

Para promedios de vectores aleatorios, el TCL es totalmente análogo:

Sean $X_1, X_2, \ldots$ vectores aleatorios p dimensionales i.i.d. con vector de medias μ y matriz de covarianzas Σ, entonces $\sqrt{n}(\bar{X}_n - \mu) \rightarrow_d \mathrm{N}_p(0, \Sigma)$, donde $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$.