IECD-325: Distribución normal multivariada

Felipe Osorio

felipe.osorio@uv.cl

Recordatorio 1:

Una variable aleatoria (uni-dimensional) Z tiene distribución normal (univariada) con media cera y varianza uno, si su pdf[1] es de la forma:

$$f(z) = (2\pi)^{-1/2} \exp(-\tfrac{1}{2} z^2), \qquad z \in \mathbb{R},$$

en cuyo caso escribimos $Z \sim \mathsf{N}(0,1)$. Más generalmente,

$$Y \sim \mathsf{N}(\mu, \sigma^2) \quad \iff \quad Y \stackrel{\mathsf{d}}{=} \mu + \sigma Z, \quad Z \sim \mathsf{N}(0,1).$$

donde $\mu \in \mathbb{R}$ es la media, mientras que $\sigma^2 \geq 0$ representa la varianza.

Recordatorio 2:

Si $Y \sim \mathsf{N}(\mu, \sigma^2)$, entonces su función característica adopta la forma:

$$\varphi(h) = \exp(ih\mu - \tfrac{1}{2}\sigma^2 h^2), \qquad h \in \mathbb{R}.$$

[1] En ocasiones escribimos $f(z) = \phi(z)$ para denotar la densidad de $Z \sim \mathsf{N}(0,1)$.

Definición 1 (Distribución normal multivariada):

Sea $\boldsymbol{X}$ vector aleatorio p-dimensional. Se dice que $\boldsymbol{X}$ tiene distribución normal multivariada con vector de medias $\boldsymbol{\mu} \in \mathbb{R}^p$ y matriz de covarianza $\boldsymbol{\Sigma} \geq \boldsymbol{0}$ si y sólo si,

$$Y = \boldsymbol{t}^\top \boldsymbol{X} \sim \mathsf{N}_1(\mu_Y, \sigma_Y^2), \qquad \text{para todo } \boldsymbol{t} \in \mathbb{R}^p,$$

y anotamos $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Observación:

En la definición anterior no se ha hecho supuestos sobre la independencia de las componentes de $\boldsymbol{X}$.

Resultado 1 (Transformación afín):

Suponga que $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere la transformación $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ donde $\boldsymbol{A} \in \mathbb{R}^{m \times p}$ con $\operatorname{rg}(\boldsymbol{A}) = m$ y $\boldsymbol{b} \in \mathbb{R}^m$. Entonces

$$\boldsymbol{Y} \sim \mathsf{N}_m(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top).$$

Demostración:

Sea $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ y note que

$$\boldsymbol{t}^\top\boldsymbol{Y} = \boldsymbol{t}^\top\boldsymbol{A}\boldsymbol{X} + \boldsymbol{t}^\top\boldsymbol{b} = (\boldsymbol{A}^\top\boldsymbol{t})^\top\boldsymbol{X} + \boldsymbol{t}^\top\boldsymbol{b} = \boldsymbol{h}^\top\boldsymbol{X} + c,$$

por la Definición 1 tenemos que $\boldsymbol{h}^\top\boldsymbol{X}$ es normal y como c es una constante, sigue que $\boldsymbol{t}^\top\boldsymbol{Y}$ tiene distribución normal multivariada.

Resultado 2 (Función característica):

Si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces la función característica de $\boldsymbol{X}$ es dada por

$$\varphi(\boldsymbol{t}) = \exp\big(i\boldsymbol{t}^\top \boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}\big).$$

Demostración:

Sea $\boldsymbol{\Sigma}$ matriz de covarianza $p \times p$ semidefinida positiva de rango r y sea $Z_1, \dots, Z_r$ variables aleatorias IID $\mathsf{N}(0, 1)$. Entonces el vector $\boldsymbol{Z} = (Z_1, \dots, Z_r)^\top$ tiene función característica

$$\begin{aligned} \varphi_Z(\boldsymbol{t}) &= \mathsf{E}\{\exp(i\boldsymbol{t}^\top \boldsymbol{Z})\} = \prod_{j=1}^r \mathsf{E}\{\exp(it_j Z_j)\} \\ &= \prod_{j=1}^r \exp\big(-\tfrac{1}{2}t_j^2\big) = \exp\big(-\tfrac{1}{2}\boldsymbol{t}^\top \boldsymbol{t}\big). \end{aligned} \tag{1}$$

Considere

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{Z},$$

donde $\boldsymbol{B} \in \mathbb{R}^{p\times r}$ con $\mathrm{rg}(\boldsymbol{B}) = r$, tal que $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$ y $\boldsymbol{\mu} \in \mathbb{R}^p$. Entonces $\boldsymbol{X}$ tiene función característica

$$\begin{aligned}
\varphi_X(\boldsymbol{t}) &= \mathsf{E}\{\exp(i\boldsymbol{t}^\top \boldsymbol{X})\} = \mathsf{E}\{\exp(i\boldsymbol{t}^\top(\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{Z}))\} \\
&= \exp(i\boldsymbol{t}^\top\boldsymbol{\mu})\,\mathsf{E}\{\exp(i\boldsymbol{t}^\top\boldsymbol{B}\boldsymbol{Z})\} = \exp(i\boldsymbol{t}^\top\boldsymbol{\mu})\varphi_Z(\boldsymbol{h}), \quad \boldsymbol{h} = \boldsymbol{B}^\top\boldsymbol{t} \\
&= \exp(i\boldsymbol{t}^\top\boldsymbol{\mu})\exp(-\tfrac{1}{2}\boldsymbol{t}^\top\boldsymbol{B}\boldsymbol{B}^\top\boldsymbol{t}) = \exp(i\boldsymbol{t}^\top\boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}).
\end{aligned}$$

Tarea:

Mostrar el Resultado 1 usando la función característica.

Resultado 3:

Si $\boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$. Entonces

$$\mathsf{E}(\boldsymbol{Z}) = \boldsymbol{0}, \qquad \mathsf{Cov}(\boldsymbol{Z}) = \boldsymbol{I}.$$

Demostración:

Para mostrar el resultado, se obtendrá

$$\mathsf{E}(\boldsymbol{Z}) = i^{-1} \frac{\partial \varphi(\boldsymbol{t})}{\partial \boldsymbol{t}}\Big|_{t=0}, \qquad \mathsf{E}(\boldsymbol{Z}\boldsymbol{Z}^\top) = i^{-2} \frac{\partial^2 \varphi(\boldsymbol{t})}{\partial \boldsymbol{t} \partial \boldsymbol{t}^\top}\Big|_{t=0}.$$

Sea $Z_1, \dots, Z_n$ variables aleatorias IID $\mathsf{N}(0, 1)$. Entonces por (1), tenemos

$$\varphi_Z(\boldsymbol{t}) = \exp(-\tfrac{1}{2}\|\boldsymbol{t}\|^2), \qquad \boldsymbol{t} \in \mathbb{R}^n.$$

Tenemos,[2]

$$\begin{aligned} \mathsf{d}\,\varphi_Z(\boldsymbol{t}) &= \exp(-\tfrac{1}{2}\|\boldsymbol{t}\|^2)\cdot(-\tfrac{1}{2}\,\mathsf{d}\,\|\boldsymbol{t}\|^2) = -\exp(-\tfrac{1}{2}\|\boldsymbol{t}\|^2)\boldsymbol{t}^\top\,\mathsf{d}\,\boldsymbol{t} \\ &= -\varphi_Z(\boldsymbol{t})\boldsymbol{t}^\top\,\mathsf{d}\,\boldsymbol{t}. \end{aligned}$$

y

$$\begin{aligned} \mathsf{d}^2\,\varphi_Z(\boldsymbol{t}) &= -\,\mathsf{d}\varphi_Z(\boldsymbol{t})\boldsymbol{t}^\top\,\mathsf{d}\boldsymbol{t} - \varphi_Z(\boldsymbol{t})(\mathsf{d}\boldsymbol{t})^\top\,\mathsf{d}\boldsymbol{t} = \varphi_Z(\boldsymbol{t})(\mathsf{d}\boldsymbol{t})^\top\boldsymbol{t}\boldsymbol{t}^\top\,\mathsf{d}\boldsymbol{t} - \varphi_Z(\boldsymbol{t})(\mathsf{d}\boldsymbol{t})^\top\,\mathsf{d}\boldsymbol{t} \\ &= \varphi_Z(\boldsymbol{t})(\mathsf{d}\boldsymbol{t})^\top(\boldsymbol{t}\boldsymbol{t}^\top - \boldsymbol{I})\,\mathsf{d}\boldsymbol{t}, \end{aligned}$$

de ahí que

$$\frac{\partial\varphi_Z(\boldsymbol{t})}{\partial\boldsymbol{t}} = -\varphi_Z(\boldsymbol{t})\boldsymbol{t}, \qquad \frac{\partial^2\varphi_Z(\boldsymbol{t})}{\partial\boldsymbol{t}\,\partial\boldsymbol{t}^\top} = \varphi_Z(\boldsymbol{t})(\boldsymbol{t}\boldsymbol{t}^\top - \boldsymbol{I}).$$

Evaluando en $\boldsymbol{t} = \boldsymbol{0}$ y ponderando de forma apropiada, sigue que

$$\mathsf{E}(\boldsymbol{Z}) = i^{-1}\,\frac{\partial\varphi_Z(\boldsymbol{t})}{\partial\boldsymbol{t}}\Big|_{t=0} = \boldsymbol{0}, \qquad \mathsf{E}(\boldsymbol{Z}\boldsymbol{Z}^\top) = i^{-2}\,\frac{\partial^2\varphi_Z(\boldsymbol{t})}{\partial\boldsymbol{t}\partial\boldsymbol{t}^\top}\Big|_{t=0} = \boldsymbol{I} = \mathsf{Cov}(\boldsymbol{Z}).$$

[2] Evidentemente $\mathsf{d}\,\|\boldsymbol{t}\|^2 = (\mathsf{d}\,\boldsymbol{t})^\top\boldsymbol{t} + \boldsymbol{t}^\top\,\mathsf{d}\,\boldsymbol{t} = 2\boldsymbol{t}^\top\,\mathsf{d}\,\boldsymbol{t}$.

Observación:

Sea

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{Z}, \qquad \boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I}),$$

con $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$ con $\boldsymbol{B}$ matriz de rango completo. Por los Resultados 1 y 3, sigue que

$$\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} \geq \boldsymbol{0}.$$

Además,

$$\begin{aligned} \mathsf{E}(\boldsymbol{X}) &= \boldsymbol{\mu} + \boldsymbol{B}\,\mathsf{E}(\boldsymbol{Z}) = \boldsymbol{\mu} \\ \mathsf{Cov}(\boldsymbol{X}) &= \mathsf{Cov}(\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{Z}) = \boldsymbol{B}\,\mathsf{Cov}(\boldsymbol{Z})\boldsymbol{B}^\top = \boldsymbol{\Sigma}. \end{aligned}$$

Resultado 4:

Si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces la distribución marginal de cualquier subconjunto de k $(< p)$ componentes de $\boldsymbol{X}$ es normal k-variada.

Demostración:

Considere la siguiente partición:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2}$$

donde $\boldsymbol{X}_1$ y $\boldsymbol{\mu}_1$ son vectores $k \times 1$ y $\boldsymbol{\Sigma}_{11}$ es $k \times k$. Aplicando el Resultado 1 con

$$\boldsymbol{A} = (\boldsymbol{I}_k, \boldsymbol{0}) \in \mathbb{R}^{k \times p} \qquad \text{y} \qquad \boldsymbol{b} = \boldsymbol{0},$$

sigue inmediatamente que $\boldsymbol{X}_1 \sim \mathsf{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Observación:

La inversa del Resultado 4, no es verdad en general.

Resultado 4:

Si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces la distribución marginal de cualquier subconjunto de k $(< p)$ componentes de $\boldsymbol{X}$ es normal k-variada.

Demostración:

Considere la siguiente partición:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2}$$

donde $\boldsymbol{X}_1$ y $\boldsymbol{\mu}_1$ son vectores $k \times 1$ y $\boldsymbol{\Sigma}_{11}$ es $k \times k$. Aplicando el Resultado 1 con

$$\boldsymbol{A} = (\boldsymbol{I}_k, \boldsymbol{0}) \in \mathbb{R}^{k \times p} \qquad \text{y} \qquad \boldsymbol{b} = \boldsymbol{0},$$

sigue inmediatamente que $\boldsymbol{X}_1 \sim \mathsf{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Observación:

La inversa del Resultado 4, no es verdad en general.

Resultado 5:

Si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\boldsymbol{X}$, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son particionadas como en la Ecuación (2). Entonces los vectores $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ son independientes sólo si $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$.

Demostración:

Note que $\operatorname{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\Sigma}_{12}$, así la independencia entre $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ implica que

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{0}.$$

Suponga ahora que $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$. Entonces la función característica

$$\begin{aligned}
\varphi_X(\boldsymbol{t}) &= \exp(i\boldsymbol{t}^\top \boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}) \\
&= \exp(i\boldsymbol{t}_1^\top \boldsymbol{\mu}_1 + i\boldsymbol{t}_2^\top \boldsymbol{\mu}_2 - \tfrac{1}{2}\boldsymbol{t}_1^\top \boldsymbol{\Sigma}_{11}\boldsymbol{t}_1 - \tfrac{1}{2}\boldsymbol{t}_2^\top \boldsymbol{\Sigma}_{22}\boldsymbol{t}_2) \\
&= \exp(i\boldsymbol{t}_1^\top \boldsymbol{\mu}_1 - \tfrac{1}{2}\boldsymbol{t}_1^\top \boldsymbol{\Sigma}_{11}\boldsymbol{t}_1) \exp(i\boldsymbol{t}_2^\top \boldsymbol{\mu}_2 - \tfrac{1}{2}\boldsymbol{t}_2^\top \boldsymbol{\Sigma}_{22}\boldsymbol{t}_2) \\
&= \varphi_{X_1}(\boldsymbol{t}_1)\varphi_{X_2}(\boldsymbol{t}_2),
\end{aligned}$$

es decir, $\boldsymbol{X}_1 \sim \mathsf{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ es independiente de $\boldsymbol{X}_2 \sim \mathsf{N}_{p-k}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

Resultado 6 (Función de densidad):

Si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\boldsymbol{\Sigma}$ es definida positiva, entonces la densidad de $\boldsymbol{X}$ asume la forma

$$f_X(\boldsymbol{x}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}, \qquad \boldsymbol{x} \in \mathbb{R}^p.$$

Demostración:

Sea $Z_1, \dots, Z_p$ variables aleatorias IID $\mathsf{N}(0,1)$. Entonces la densidad conjunta de $\boldsymbol{Z} = (Z_1, \dots, Z_p)^\top$ es

$$f_Z(\boldsymbol{z}) = \prod_{i=1}^p (2\pi)^{-1/2} \exp(-z_i^2/2) = (2\pi)^{-p/2} \exp(-\tfrac{1}{2}\|\boldsymbol{z}\|^2).$$

Considere $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{Z}$ con $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$, con $\boldsymbol{B}$ matriz de rango completo. Entonces, tenemos la transformación inversa

$$\boldsymbol{Z} = \boldsymbol{g}^{-1}(\boldsymbol{X}) = \boldsymbol{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}),$$

De este modo,

$$\mathsf{d}\boldsymbol{Z} = \mathsf{d}\boldsymbol{g}^{-1}(\boldsymbol{X}) = \boldsymbol{B}^{-1}\,\mathsf{d}\boldsymbol{X},$$

con matriz Jacobiana $\mathsf{D}\,\boldsymbol{g}^{-1}(\boldsymbol{X}) = \boldsymbol{B}^{-1}$, como

$$|\,\mathsf{D}\,\boldsymbol{g}^{-1}(\boldsymbol{X})|_{+} = |\boldsymbol{B}|^{-1} = |\boldsymbol{B}\boldsymbol{B}^{\top}|^{-1/2},$$

obtenemos

$$\begin{aligned} f_X(\boldsymbol{x}) &= |\,\mathsf{D}\,\boldsymbol{g}^{-1}(\boldsymbol{x})|_{+} f_Z(\boldsymbol{g}^{-1}(\boldsymbol{x})) \\ &= (2\pi)^{-p/2}|\boldsymbol{B}\boldsymbol{B}^{\top}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\boldsymbol{B}^{-\top}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}, \end{aligned}$$

notando que $\boldsymbol{\Sigma}^{-1} = (\boldsymbol{B}\boldsymbol{B}^{\top}) = \boldsymbol{B}^{-\top}\boldsymbol{B}^{-1}$ sigue el resultado deseado.

Ejemplo:

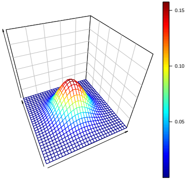
Sea $\boldsymbol{X} \sim \mathsf{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma})$ donde

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad -1 < \rho < 1.$$

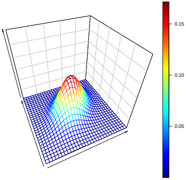
En cuyo caso, la función de densidad es dada por:

$$f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}(x_1^2 + x_2^2 - 2\rho x_1 x_2)\Big\}.$$

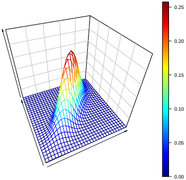
A continuación se presenta la función de densidad para los casos $\rho = 0.0, 0.4$ y 0.8.



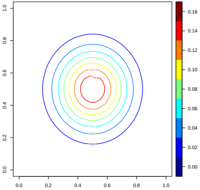
(a) $\rho = 0.0$



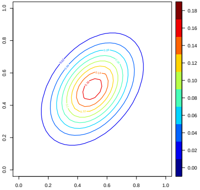
(b) $\rho = 0.4$



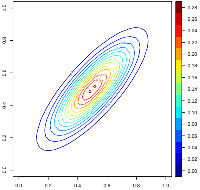
(c) $\rho = 0.8$



(d) contornos



(e) contornos



(f) contornos

Observación:

Es fácil notar que la función de densidad es constante sobre el elipsoide

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \lambda,$$

en $\mathbb{R}^p$ para todo $\lambda > 0$. Este elipsoide tiene centro $\boldsymbol{\mu}$, mientras que $\boldsymbol{\Sigma}$ determina su forma y orientación. Además, la variable aleatoria

$$(\boldsymbol{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) = \boldsymbol{Z}^\top \boldsymbol{Z} = \sum_{i=1}^{p} Z_i^2, \qquad (3)$$

sigue una distribución chi-cuadrado con p grados de libertad y la cantidad

$$D = \{(\boldsymbol{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^{1/2},$$

se conoce como distancia de Mahalanobis[3] de $\boldsymbol{X}$ a $\boldsymbol{\mu}$.

[3] En ocasiones, decimos que $D^2 = (\boldsymbol{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$ es la 'distancia' de Mahalanobis.

Resultado 7:

Sea $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y particione $\boldsymbol{X}$, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ como:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

donde $\boldsymbol{X}_1$ y $\boldsymbol{\mu}_1$ son vectores $k \times 1$, mientras que $\boldsymbol{\Sigma}_{11}$ es matriz $k \times k$. Sea $\boldsymbol{\Sigma}_{22}^{-}$ una inversa generalizada de $\boldsymbol{\Sigma}_{22}$, esto es, una matriz que satisface

$$\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{22},$$

y sea $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\Sigma}_{21}$. Entonces

(a) $\boldsymbol{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{X}_2 \sim \mathsf{N}_k(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11\cdot 2})$ y es independiente de $\boldsymbol{X}_2$.

(b) La distribución condicional

$$(\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2) \sim \mathsf{N}_k(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11\cdot 2}).$$

Demostración:

Considere la transformación lineal

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_k & -\boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I}_{p-k} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} = \boldsymbol{C}\boldsymbol{X},$$

sigue que $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{C}\boldsymbol{\mu}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top)$, donde

$$\begin{aligned} \boldsymbol{C}\boldsymbol{\mu} &= \begin{pmatrix} \boldsymbol{I}_k & -\boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I}_{p-k} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{B}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix} \\ \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top &= \begin{pmatrix} \boldsymbol{I}_k & -\boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I}_{p-k} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ -\boldsymbol{B}^\top & \boldsymbol{I}_{p-k} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{B}\boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{12}\boldsymbol{B}^\top + \boldsymbol{B}\boldsymbol{\Sigma}_{22}\boldsymbol{B}^\top & \boldsymbol{\Sigma}_{12} - \boldsymbol{B}\boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{22}\boldsymbol{B}^\top & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \end{aligned}$$

De este modo, nuestro interés es escoger $\boldsymbol{\Sigma}_{12} - \boldsymbol{B}\boldsymbol{\Sigma}_{22} = \boldsymbol{0}$. Es decir, $\boldsymbol{\Sigma}_{12} = \boldsymbol{B}\boldsymbol{\Sigma}_{22}$.

Por otro lado, notando que

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\Sigma}_{22} = \boldsymbol{B}\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\Sigma}_{22} = \boldsymbol{B}\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{12},$$

sigue que $\boldsymbol{\Sigma}_{12}\boldsymbol{B}^\top = \boldsymbol{B}\boldsymbol{\Sigma}_{22}\boldsymbol{B}^\top$ (y análogamente $\boldsymbol{B}\boldsymbol{\Sigma}_{21} = \boldsymbol{B}\boldsymbol{\Sigma}_{22}\boldsymbol{B}^\top$).

Esto es, si $\boldsymbol{B}$ es escogida como $\boldsymbol{B} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}$, entonces $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ son independientes con distribución conjunta

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{X}_2 \\ \boldsymbol{X}_2 \end{pmatrix} \sim \mathsf{N}_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right).$$

Esto muestra la parte (a).

Para notar la parte (b), note que las densidades de $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ están dadas por

$$g(\boldsymbol{y}_1;\boldsymbol{\delta}_{1\cdot 2},\boldsymbol{\Sigma}_{11\cdot 2}) = |2\pi\boldsymbol{\Sigma}_{11\cdot 2}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{y}_1-\boldsymbol{\delta}_{1\cdot 2})^\top\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{y}_1-\boldsymbol{\delta}_{1\cdot 2})\}$$
$$f_2(\boldsymbol{y}_2;\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22}) = |2\pi\boldsymbol{\Sigma}_{22}|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{y}_2-\boldsymbol{\mu}_2)^\top\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2-\boldsymbol{\mu}_2)\},$$

donde $\boldsymbol{\delta}_{1\cdot 2} = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\mu}_2$. La densidad conjunta para $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top,\boldsymbol{Y}_2^\top)^\top$ adopta la forma

$$f(\boldsymbol{y}_1,\boldsymbol{y}_2;\boldsymbol{\mu},\boldsymbol{\Sigma}) = g(\boldsymbol{y}_1;\boldsymbol{\delta}_{1\cdot 2},\boldsymbol{\Sigma}_{11\cdot 2})\, f_2(\boldsymbol{y}_2;\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22}).$$

Como

$$f(\boldsymbol{x}_1,\boldsymbol{x}_2;\boldsymbol{\mu},\boldsymbol{\Sigma}) = f_{1|2}(\boldsymbol{x}_1;\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{x}_2)\, f_2(\boldsymbol{x}_2;\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22}),$$

entonces, la densidad condicional de $\boldsymbol{X}_1$ dado $\boldsymbol{X}_2 = \boldsymbol{x}_2$ debe ser $g(\boldsymbol{y}_1;\boldsymbol{\delta}_{1\cdot 2},\boldsymbol{\Sigma}_{11\cdot 2})$.

Además, es fácil notar que la forma cuadrática

$$\begin{aligned} q(\boldsymbol{y}_1; \boldsymbol{\mu}_{1\cdot 2}, \boldsymbol{\Sigma}_{11\cdot 2}) &= (\boldsymbol{y}_1 - \boldsymbol{\delta}_{1\cdot 2})^\top \boldsymbol{\Sigma}_{11\cdot 2}^{-1} (\boldsymbol{y}_1 - \boldsymbol{\delta}_{1\cdot 2}) \\ &= (\boldsymbol{x}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{x}_2 - \boldsymbol{\delta}_{1\cdot 2})^\top \boldsymbol{\Sigma}_{11\cdot 2}^{-1} (\boldsymbol{x}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{x}_2 - \boldsymbol{\delta}_{1\cdot 2}) \\ &= (\boldsymbol{x}_1 - \boldsymbol{\mu}_{1\cdot 2})^\top \boldsymbol{\Sigma}_{11\cdot 2}^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_{1\cdot 2}), \end{aligned}$$

donde

$$\boldsymbol{\mu}_{1\cdot 2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),$$

lo que muestra el resultado.

Observación:

La esperanza de la distribución condicional de $\boldsymbol{X}_1$ dado $\boldsymbol{X}_2$, es decir

$$\mathsf{E}(\boldsymbol{X}_1|\boldsymbol{X}_2 = \boldsymbol{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),$$

se denomina función de regresión de $\boldsymbol{X}_1$ sobre $\boldsymbol{X}_2$ con coeficientes de regresión $\boldsymbol{B} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}$. Esta es una función lineal de $\boldsymbol{X}_2$ y la matriz de covarianza $\boldsymbol{\Sigma}_{11\cdot 2}$ no depende de $\boldsymbol{X}_2$.

Resultado 8:

Sea $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere $\boldsymbol{Y}_1 = \boldsymbol{A}_1\boldsymbol{X}$, $\boldsymbol{Y}_2 = \boldsymbol{A}_2\boldsymbol{X}$ dos funciones lineales del vector aleatorio $\boldsymbol{X}$. La covarianza entre $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ es dada por

$$\mathsf{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \boldsymbol{A}_1 \mathsf{Cov}(\boldsymbol{X}, \boldsymbol{X})\boldsymbol{A}_2^\top = \boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2^\top$$

Observación:

Por el resultado anterior es evidente que $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ serán independientes si y sólo si

$$\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2^\top = \boldsymbol{0}.$$

Ejemplo:

Considere $X_1, \dots, X_n$ una muestra aleatoria desde $\mathsf{N}(\mu, \sigma^2)$ y sea

$$\boldsymbol{Z} = (Z_1, \dots, Z_n)^\top,$$

el vector de datos centrados con $Z_i = X_i - \overline{X}$, $i = 1, \dots, n$, donde $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Podemos escribir

$$\overline{X} = \frac{1}{n}\boldsymbol{1}^\top \boldsymbol{X}, \qquad \boldsymbol{Z} = \boldsymbol{C}\boldsymbol{X},$$

con $\boldsymbol{C} = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top$ la matriz de centrado. Tenemos que $\boldsymbol{X} \sim \mathsf{N}_n(\mu\boldsymbol{1}, \sigma^2\boldsymbol{I}_n)$ y $\overline{X}$ con $\boldsymbol{Z}$ son independientes pues $\boldsymbol{C}\boldsymbol{1} = \boldsymbol{0}$.

Ejemplo:

Sea $\boldsymbol{X} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$ y considere las transformaciones

$$\boldsymbol{Y}_1 = \boldsymbol{A}\boldsymbol{X}, \qquad \boldsymbol{Y}_2 = (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})^\top \boldsymbol{X}.$$

De este modo

$$\mathsf{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \mathsf{Cov}(\boldsymbol{A}\boldsymbol{X}, (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})^\top \boldsymbol{X}) = \sigma^2 \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A}) = \mathbf{0},$$

pues $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$ y $\boldsymbol{Y}_1$ con $\boldsymbol{Y}_2$ son independientes.