

MAT-266: Distribución normal multivariada

Felipe Osorio

`fosorios.mat.utfsm.cl`

Departamento de Matemática, UTFSM



Recordatorio 1:

Una variable aleatoria (uni-dimensional) Z tiene **distribución normal** (univariada) con media cero y varianza uno, si su pdf¹ es de la forma:

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

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

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

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

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

Recordatorio 2:

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

$$\varphi(t) = \exp(it\mu - \frac{1}{2}\sigma^2 t^2), \quad t \in \mathbb{R}.$$

Recordatorio 3:

Sea Z_1, \dots, Z_n variables aleatorias **independientes** cada una con distribución $N(0, 1)$ y considere el vector aleatorio $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$. Entonces,

$$E(\mathbf{Z}) = \mathbf{0}, \quad \text{Cov}(\mathbf{Z}) = \mathbf{I}.$$



Definición 1 (Distribución normal multivariada)

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

$$Y = \mathbf{t}^\top \mathbf{X} \sim N_1(\mu_Y, \sigma_Y^2), \quad \text{para todo } \mathbf{t} \in \mathbb{R}^p,$$

y anotamos $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Observación:

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



Resultado 1 (Transformación afín)

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

$$\mathbf{Y} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top).$$

Demostración:

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

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

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



Resultado 2 (Función característica):

Si $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces la **función característica** de \mathbf{X} es dada por

$$\varphi(\mathbf{t}) = \exp\left(i\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}\right).$$

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 $N(0, 1)$. Entonces el vector $\mathbf{Z} = (Z_1, \dots, Z_r)^\top$ tiene función característica

$$\begin{aligned}\varphi_{\mathbf{Z}}(\mathbf{t}) &= \mathbb{E}\{\exp(i\mathbf{t}^\top \mathbf{Z})\} = \prod_{j=1}^r \mathbb{E}\{\exp(it_j Z_j)\} \\ &= \prod_{j=1}^r \exp\left(-\frac{1}{2}t_j^2\right) = \exp\left(-\frac{1}{2}\mathbf{t}^\top \mathbf{t}\right).\end{aligned}\tag{1}$$



Considere

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

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

$$\begin{aligned}\varphi_{\mathbf{X}}(\mathbf{t}) &= \mathbb{E}\{\exp(i\mathbf{t}^\top \mathbf{X})\} = \mathbb{E}\{\exp(i\mathbf{t}^\top (\boldsymbol{\mu} + \mathbf{B}\mathbf{Z}))\} \\ &= \exp(i\mathbf{t}^\top \boldsymbol{\mu}) \mathbb{E}\{\exp(i\mathbf{t}^\top \mathbf{B}\mathbf{Z})\} = \exp(i\mathbf{t}^\top \boldsymbol{\mu}) \varphi_{\mathbf{Z}}(\mathbf{h}), \quad \mathbf{h} = \mathbf{B}^\top \mathbf{t} \\ &= \exp(i\mathbf{t}^\top \boldsymbol{\mu}) \exp(-\frac{1}{2}\mathbf{t}^\top \mathbf{B}\mathbf{B}^\top \mathbf{t}) = \exp(i\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}).\end{aligned}$$

Tarea:

Mostrar el [Resultado 1](#) usando la función característica.



Resultado 3:

Si $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I})$. Entonces

$$E(\mathbf{Z}) = \mathbf{0}, \quad \text{Cov}(\mathbf{Z}) = \mathbf{I}.$$

Demostración:

Para mostrar el resultado, se obtendrá

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

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

$$\varphi_Z(\mathbf{t}) = \exp(-\frac{1}{2} \|\mathbf{t}\|^2), \quad \mathbf{t} \in \mathbb{R}^n.$$



Distribución normal multivariada

Tenemos,²

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

y

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

de ahí que

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

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

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

²Evidentemente $d\|\mathbf{t}\|^2 = (d\mathbf{t})^\top \mathbf{t} + \mathbf{t}^\top d\mathbf{t} = 2\mathbf{t}^\top d\mathbf{t}$.



Observación:

Sea

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{B}\mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}),$$

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

$$\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} \geq \mathbf{0}.$$

Además,

$$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu} + \mathbf{B} \mathbb{E}(\mathbf{Z}) = \boldsymbol{\mu}$$

$$\text{Cov}(\mathbf{X}) = \text{Cov}(\boldsymbol{\mu} + \mathbf{B}\mathbf{Z}) = \mathbf{B} \text{Cov}(\mathbf{Z}) \mathbf{B}^\top = \boldsymbol{\Sigma}.$$



Resultado 4:

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

Demostración:

Considere la siguiente partición:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (2)$$

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

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

sigue inmediatamente que $\mathbf{X}_1 \sim N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Observación:

La inversa del [Resultado 4](#), **no** es verdad en general.



Resultado 4:

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

Demostración:

Considere la siguiente partición:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (2)$$

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

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

sigue inmediatamente que $\mathbf{X}_1 \sim N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Observación:

La inversa del [Resultado 4](#), **no** es verdad en general.



Resultado 5:

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

Demostración:

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

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

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

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

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



Resultado 6 (Función de densidad):

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

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

Demostración:

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

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^p (2\pi)^{-1/2} \exp(-z_i^2/2) = (2\pi)^{-p/2} \exp(-\frac{1}{2}\|\mathbf{z}\|^2).$$

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

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



De este modo,

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

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

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

obtenemos

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

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



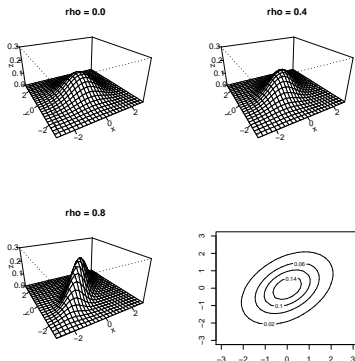
Distribución normal multivariada

Ejemplo:

Sea $\mathbf{X} \sim N_2(\mathbf{0}, \Sigma)$ donde

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

La siguiente figura presenta la función de densidad para los casos $\rho = 0.0, 0.4$ y 0.8 .



Observación:

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

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

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

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

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

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

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

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

Resultado 7:

Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y particione \mathbf{X} , $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ como:

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

donde \mathbf{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) $\mathbf{X}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^- \mathbf{X}_2 \sim N_k(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^- \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11 \cdot 2})$ y es independiente de \mathbf{X}_2 .
- (b) La distribución condicional

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



Demostración:

Considere la transformación lineal

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} I_k & -B \\ \mathbf{0} & I_{p-k} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = CX,$$

sigue que $Y \sim N_p(C\mu, C\Sigma C^\top)$, donde

$$\begin{aligned} C\mu &= \begin{pmatrix} I_k & -B \\ \mathbf{0} & I_{p-k} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 - B\mu_2 \\ \mu_2 \end{pmatrix} \\ C\Sigma C^\top &= \begin{pmatrix} I_k & -B \\ \mathbf{0} & I_{p-k} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_k & \mathbf{0} \\ -B^\top & I_{p-k} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} - B\Sigma_{21} - \Sigma_{12}B^\top + B\Sigma_{22}B^\top & \Sigma_{12} - B\Sigma_{22} \\ \Sigma_{21} - \Sigma_{22}B^\top & \Sigma_{22} \end{pmatrix}. \end{aligned}$$

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



Por otro lado, notando que

$$\Sigma_{12}\Sigma_{22}^{-}\Sigma_{22} = B\Sigma_{22}\Sigma_{22}^{-}\Sigma_{22} = B\Sigma_{22} = \Sigma_{12},$$

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

Esto es, si B es escogida como $B = \Sigma_{12}\Sigma_{22}^{-}$, entonces Y_1 y Y_2 son independientes con distribución conjunta

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2 \\ X_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-}\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right).$$

Esto muestra la parte (a).



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

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

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

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

Como

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

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



Además, es fácil notar que la forma cuadrática

$$\begin{aligned} q(\mathbf{y}_1; \boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}) &= (\mathbf{y}_1 - \boldsymbol{\delta}_{1.2})^\top \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{y}_1 - \boldsymbol{\delta}_{1.2}) \\ &= (\mathbf{x}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \mathbf{x}_2 - \boldsymbol{\delta}_{1.2})^\top \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{x}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \mathbf{x}_2 - \boldsymbol{\delta}_{1.2}) \\ &= (\mathbf{x}_1 - \boldsymbol{\mu}_{1.2})^\top \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_{1.2}), \end{aligned}$$

donde

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2),$$

lo que muestra el resultado.



Observación:

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

$$E(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2),$$

se denomina **función de regresión** de \mathbf{X}_1 sobre \mathbf{X}_2 con coeficientes de regresión $\mathbf{B} = \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}$. Esta es una función lineal de \mathbf{X}_2 y la matriz de covarianza $\boldsymbol{\Sigma}_{11 \cdot 2}$ no depende de \mathbf{X}_2 .

Resultado 8:

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

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

Observación:

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

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



Ejemplo:

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

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

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

$$\bar{X} = \frac{1}{n} \mathbf{1}^\top \mathbf{X}, \quad \mathbf{Z} = \mathbf{C} \mathbf{X},$$

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



Ejemplo:

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

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

De este modo

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

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

