

Preliminares

1.1. Vectores Aleatorios

El propósito de esta sección es introducir algunas propiedades elementales de vectores aleatorios útiles a lo largo de este curso. Se asume que el lector es familiar con el concepto de variable aleatoria unidimensional.

Un vector aleatorio n -dimensional \mathbf{X} es una función (medible) desde el espacio de probabilidad Ω a \mathbb{R}^n , esto es

$$\mathbf{X} : \Omega \rightarrow \mathbb{R}^n.$$

Por convención asumiremos que el vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)^\top$ es un vector columna.

DEFINICIÓN 1.1 (Función de distribución). Para \mathbf{X} distribuido en \mathbb{R}^n , la *función de distribución* de \mathbf{X} es una función $F : \mathbb{R}^n \rightarrow [0, 1]$, tal que

$$F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \quad (1.1)$$

y denotamos $\mathbf{X} \sim F$ o $\mathbf{X} \sim F_X$.

La función en (1.1) debe ser entendida como

$$F(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n),$$

que corresponde a la probabilidad del evento $\bigcap_{k=1}^n \{X_k \leq x_k\}$.

La función de distribución acumulada tiene las siguientes propiedades:

- (a) $F(\mathbf{x})$ es función monótona creciente y continua a la derecha en cada uno de los componentes de \mathbf{X} ,
- (b) $0 \leq F(\mathbf{x}) \leq 1$,
- (c) $F(-\infty, x_2, \dots, x_n) = \dots = F(x_1, \dots, x_{n-1}, -\infty) = 0$,
- (d) $F(+\infty, \dots, +\infty) = 1$.

Sea F la función de distribución del vector aleatorio \mathbf{X} . Entonces, existe una función no-negativa f tal que

$$F(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} f(\mathbf{u}) \, d\mathbf{u}, \quad \mathbf{x} \in \mathbb{R}^n,$$

en este caso decimos que \mathbf{X} es un vector aleatorio continuo con *función de densidad* f . Por el teorema fundamental del Cálculo, tenemos que

$$f(\mathbf{x}) = \frac{\partial^n F(\mathbf{x})}{\partial x_1 \cdots \partial x_n}.$$

Además, considere $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, para \mathbf{x}, \mathbf{y} vectores en $\bar{\mathbb{R}}^n$, entonces

$$\mathbf{x} \leq \mathbf{y} \quad \text{esto es,} \quad x_i \leq y_i, i = 1, \dots, n.$$

Esto permite definir un rectángulo n -dimensional en \mathbb{R}^n como

$$I = (\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} < \mathbf{x} \leq \mathbf{b}\}$$

para todo $\mathbf{a}, \mathbf{b} \in \bar{\mathbb{R}}^n$. Entonces, también por el teorema fundamental del Cálculo, tenemos que si

$$f(\mathbf{x}) = \frac{\partial^n F(\mathbf{x})}{\partial x_1 \cdots \partial x_n}.$$

existe y es continua (casi en toda parte) sobre un rectángulo I , entonces

$$P(\mathbf{x} \in A) = \int_A f(\mathbf{x}) d\mathbf{x}, \quad \forall A \subset I.$$

Naturalmente la función de densidad debe satisfacer

$$\int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} = 1.$$

Considere el vector aleatorio n -dimensional \mathbf{X} particionado como $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ donde \mathbf{X}_1 y \mathbf{X}_2 son vectores $n_1 \times 1$ y $n_2 \times 1$, respectivamente, con $n = n_1 + n_2$. Tenemos que $\mathbf{X}_i \sim F_i$, $i = 1, 2$, de este modo \mathbf{X} se denomina la *conjunta* de $\mathbf{X}_1, \mathbf{X}_2$ mientras que los \mathbf{X}_1 y \mathbf{X}_2 son llamados *marginales* de \mathbf{X} .

Note que, las funciones de distribución marginal pueden ser recuperadas desde la distribución conjunta mediante

$$F_1(\mathbf{s}) = F(\mathbf{s}, \infty), \quad F_2(\mathbf{t}) = F(\infty, \mathbf{t}), \quad \forall \mathbf{s} \in \mathbb{R}^{n_1}, \mathbf{t} \in \mathbb{R}^{n_2}.$$

Cuando \mathbf{X} es absolutamente continua con función de densidad $f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2)$, entonces la función de densidad de \mathbf{X}_i también es absolutamente continua y puede ser obtenida como

$$f_1(\mathbf{s}) = \int_{\mathbb{R}^{n_2}} f(\mathbf{s}, \mathbf{u}) d\mathbf{u}, \quad f_2(\mathbf{t}) = \int_{\mathbb{R}^{n_1}} f(\mathbf{u}, \mathbf{t}) d\mathbf{u}, \quad \forall \mathbf{s} \in \mathbb{R}^{n_1}, \mathbf{t} \in \mathbb{R}^{n_2},$$

el resultado anterior es análogo para el caso de distribuciones discretas. Si \mathbf{X} es absolutamente continuo y $f_1(\mathbf{x}_1) > 0$, entonces la *densidad condicional* de \mathbf{X}_2 dado $\mathbf{X}_1 = \mathbf{x}_1$ es

$$f_{X_2|X_1=\mathbf{x}_1}(\mathbf{u}) = \frac{f_X(\mathbf{x}_1, \mathbf{u})}{f_1(\mathbf{x}_1)},$$

con función de distribución de \mathbf{X}_2 condicional a $\mathbf{X}_1 = \mathbf{x}_1$ dada por

$$F_{X_2|X_1=\mathbf{x}_1}(\mathbf{u}) = \int_{-\infty}^{\mathbf{u}} f_{X_2|X_1=\mathbf{x}_1}(\mathbf{t}) d\mathbf{t},$$

tenemos además que

$$f_{X_2|X_1=\mathbf{x}_1}(\mathbf{u}) = \frac{f_X(\mathbf{x}_1, \mathbf{u})}{\int_{\mathbb{R}^{n_2}} f_X(\mathbf{x}_1, \mathbf{t}) d\mathbf{t}}.$$

1.2. Operadores de esperanza y covarianza

Considere $\mathbf{X} = (X_1, \dots, X_n)^\top$ vector aleatorio n -dimensional con función de densidad f . Entonces la esperanza de cualquier función g de \mathbf{X} está dada por

$$\mathbb{E}(g(\mathbf{X})) = \int_{\mathbb{R}^n} g(\mathbf{t}) f(\mathbf{t}) d\mathbf{t},$$

siempre que la integral (n -dimensional) exista.

Más generalmente, sea $\mathbf{Z} = (Z_{ij})$ una función matricial $m \times n$, entonces podemos definir el operador de esperanza de una matriz aleatoria como

$$\mathbb{E}(\mathbf{Z}(\mathbf{X})) = (\mathbb{E}(Z_{ij})), \quad Z_{ij} = Z_{ij}(\mathbf{X}). \quad (1.2)$$

De la definición en (1.2) se desprenden una serie de resultados útiles con relación al operador de esperanza. Por ejemplo, sea $\mathbf{A} = (a_{ij})$ una matriz de constantes, entonces

$$\mathbb{E}(\mathbf{A}) = \mathbf{A}.$$

RESULTADO 1.2. Sea $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ y $\mathbf{C} = (c_{ij})$ matrices de constantes $l \times m$, $n \times p$ y $l \times p$, respectivamente. Entonces

$$\mathbb{E}(\mathbf{AZB} + \mathbf{C}) = \mathbf{A} \mathbb{E}(\mathbf{Z}) \mathbf{B} + \mathbf{C}.$$

DEMOSTRACIÓN. Sea $\mathbf{Y} = \mathbf{AZB} + \mathbf{C}$, entonces

$$Y_{ij} = \sum_{r=1}^m \sum_{s=1}^n a_{ir} Z_{rs} b_{sj} + c_{ij},$$

de este modo

$$\begin{aligned} \mathbb{E}(\mathbf{AZB} + \mathbf{C}) &= (\mathbb{E}(Y_{ij})) = \left(\sum_{r=1}^m \sum_{s=1}^n a_{ir} \mathbb{E}(Z_{rs}) b_{sj} + c_{ij} \right) \\ &= \mathbf{A} \mathbb{E}(\mathbf{Z}) \mathbf{B} + \mathbf{C}. \end{aligned}$$

□

Un caso particular importante corresponde a la esperanza de una transformación lineal. Considere el vector aleatorio n -dimensional, $\mathbf{Y} = \mathbf{AX}$, donde \mathbf{X} es vector aleatorio $m \times 1$, entonces $\mathbb{E}(\mathbf{AX}) = \mathbf{A} \mathbb{E}(\mathbf{X})$. Esta propiedad puede ser extendida para sumas de vectores aleatorios, como

$$\mathbb{E} \left(\sum_i \mathbf{A}_i \mathbf{X}_i \right) = \sum_i \mathbf{A}_i \mathbb{E}(\mathbf{X}_i),$$

de manera similar tenemos que

$$\mathbb{E} \left(\sum_i \alpha_i \mathbf{Z}_i \right) = \sum_i \alpha_i \mathbb{E}(\mathbf{Z}_i),$$

donde α_i son constantes y los \mathbf{Z}_i son matrices aleatorias.

DEFINICIÓN 1.3 (Matriz de covarianza). Sean \mathbf{X} e \mathbf{Y} vectores aleatorios m y n -dimensionales, respectivamente. Se define la *matriz de covarianza* entre \mathbf{X} e \mathbf{Y} como la matriz $m \times n$,

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = (\text{Cov}(X_i, Y_j)).$$

Podemos apreciar, a partir de la definición de covarianza que

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}\{(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^\top\}.$$

En efecto, sean $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X})$ y $\boldsymbol{\eta} = \mathbb{E}(\mathbf{Y})$. Entonces,

$$\begin{aligned}\text{Cov}(\mathbf{X}, \mathbf{Y}) &= (\text{Cov}(X_i, Y_j)) = (\mathbb{E}(X_i - \mu_i)(Y_j - \eta_j)) \\ &= \mathbb{E}([(X_i - \mu_i)(Y_j - \eta_j)]) = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\eta})^\top].\end{aligned}$$

Tenemos además el siguiente resultado

$$\begin{aligned}\text{Cov}(\mathbf{X}, \mathbf{Y}) &= \mathbb{E}\{(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^\top\} \\ &= \mathbb{E}(\mathbf{X}\mathbf{Y}^\top - \mathbb{E}(\mathbf{X})\mathbf{Y}^\top - \mathbf{X}\mathbb{E}^\top(\mathbf{Y}) + \mathbb{E}(\mathbf{X})\mathbb{E}^\top(\mathbf{Y})) \\ &= \mathbb{E}(\mathbf{X}\mathbf{Y}^\top) - \mathbb{E}(\mathbf{X})\mathbb{E}^\top(\mathbf{Y}).\end{aligned}$$

Se define la *matriz de dispersión (varianza)*, como $\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X})$. De este modo, tenemos

$$\text{Cov}(\mathbf{X}) = (\text{Cov}(X_i, X_j)) = \mathbb{E}\{(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^\top\},$$

y, de la misma manera que para el caso de la matriz de covarianza,

$$\text{Cov}(\mathbf{X}) = \mathbb{E}(\mathbf{X}\mathbf{X}^\top) - \mathbb{E}(\mathbf{X})\mathbb{E}^\top(\mathbf{X}).$$

EJEMPLO 1.4. Sea \mathbf{a} vector de constantes $n \times 1$, entonces

$$\text{Cov}(\mathbf{X} - \mathbf{a}) = \text{Cov}(\mathbf{X}).$$

En efecto, note que

$$\mathbf{X} - \mathbf{a} - \mathbb{E}(\mathbf{X} - \mathbf{a}) = \mathbf{X} - \mathbb{E}(\mathbf{X}),$$

por tanto, tenemos

$$\text{Cov}(\mathbf{X} - \mathbf{a}, \mathbf{X} - \mathbf{a}) = \text{Cov}(\mathbf{X}, \mathbf{X})$$

RESULTADO 1.5. Si \mathbf{X} e \mathbf{Y} son vectores aleatorios m y n -dimensionales, respectivamente y $\mathbf{A} \in \mathbb{R}^{l \times m}$, $\mathbf{B} \in \mathbb{R}^{p \times n}$, entonces

$$\text{Cov}(\mathbf{A}\mathbf{X}, \mathbf{B}\mathbf{Y}) = \mathbf{A} \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^\top.$$

DEMOSTRACIÓN. Sean $\mathbf{U} = \mathbf{A}\mathbf{X}$ y $\mathbf{V} = \mathbf{B}\mathbf{Y}$, entonces

$$\begin{aligned}\text{Cov}(\mathbf{A}\mathbf{X}, \mathbf{B}\mathbf{Y}) &= \text{Cov}(\mathbf{U}, \mathbf{V}) = \mathbb{E}\{(\mathbf{U} - \mathbb{E}(\mathbf{U}))(\mathbf{V} - \mathbb{E}(\mathbf{V}))^\top\} \\ &= \mathbb{E}\{(\mathbf{A}\mathbf{X} - \mathbf{A}\mathbb{E}(\mathbf{X}))(\mathbf{B}\mathbf{Y} - \mathbf{B}\mathbb{E}(\mathbf{Y}))^\top\} \\ &= \mathbb{E}\{\mathbf{A}(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^\top \mathbf{B}^\top\} \\ &= \mathbf{A} \mathbb{E}\{(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^\top\} \mathbf{B}^\top \\ &= \mathbf{A} \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^\top.\end{aligned}$$

□

Tenemos el siguiente caso particular,

$$\text{Cov}(\mathbf{A}\mathbf{X}) = \text{Cov}(\mathbf{A}\mathbf{X}, \mathbf{A}\mathbf{X}) = \mathbf{A} \text{Cov}(\mathbf{X}, \mathbf{X}) \mathbf{A}^\top = \mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}^\top.$$

EJEMPLO 1.6. Considere \mathbf{X} , \mathbf{Y} , \mathbf{U} y \mathbf{V} vectores aleatorios n -dimensionales y \mathbf{A} , \mathbf{B} , \mathbf{C} y \mathbf{D} matrices de órdenes apropiados, entonces

$$\begin{aligned}\text{Cov}(\mathbf{AX} + \mathbf{BY}, \mathbf{CU} + \mathbf{DV}) &= \mathbf{A} \text{Cov}(\mathbf{X}, \mathbf{U}) \mathbf{C}^\top + \mathbf{A} \text{Cov}(\mathbf{X}, \mathbf{V}) \mathbf{D}^\top \\ &\quad + \mathbf{B} \text{Cov}(\mathbf{Y}, \mathbf{U}) \mathbf{C}^\top + \mathbf{B} \text{Cov}(\mathbf{Y}, \mathbf{V}) \mathbf{D}^\top.\end{aligned}$$

tomando $\mathbf{U} = \mathbf{X}$, $\mathbf{V} = \mathbf{Y}$, $\mathbf{C} = \mathbf{A}$ y $\mathbf{D} = \mathbf{B}$, tenemos

$$\begin{aligned}\text{Cov}(\mathbf{AX} + \mathbf{BY}) &= \text{Cov}(\mathbf{AX} + \mathbf{BY}, \mathbf{AX} + \mathbf{BY}) \\ &= \mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}^\top + \mathbf{A} \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^\top \\ &\quad + \mathbf{B} \text{Cov}(\mathbf{Y}, \mathbf{X}) \mathbf{A}^\top + \mathbf{B} \text{Cov}(\mathbf{Y}) \mathbf{B}^\top.\end{aligned}$$

RESULTADO 1.7. *Toda matriz de dispersión es simétrica y semidefinida positiva*

DEMOSTRACIÓN. La simetría de la matriz de dispersión es obvia. Para mostrar que $\text{Cov}(\mathbf{X})$ es semidefinida positiva, sea $\mathbf{Z} = \mathbf{X} - \mathbf{E}(\mathbf{X})$, y considere la variable aleatoria $Y = \mathbf{a}^\top \mathbf{Z}$, para $\mathbf{a} \in \mathbb{R}^n$ un vector arbitrario. Entonces,

$$\begin{aligned}\mathbf{a}^\top \text{Cov}(\mathbf{X}) \mathbf{a} &= \mathbf{a}^\top \mathbf{E}(\mathbf{X} - \mathbf{E}(\mathbf{X}))(\mathbf{X} - \mathbf{E}(\mathbf{X}))^\top \mathbf{a} \\ &= \mathbf{E}(\mathbf{a}^\top (\mathbf{X} - \mathbf{E}(\mathbf{X}))(\mathbf{X} - \mathbf{E}(\mathbf{X}))^\top \mathbf{a}) \\ &= \mathbf{E}(\mathbf{a}^\top \mathbf{Z} \mathbf{Z}^\top \mathbf{a}) = \mathbf{E}(Y^2) \geq 0\end{aligned}$$

y por tanto, $\text{Cov}(\mathbf{X})$ es semidefinida positiva.

Ahora, suponga que $\text{Cov}(\mathbf{X})$ es semidefinida positiva de rango r ($r \leq n$). Luego $\text{Cov}(\mathbf{X}) = \mathbf{B} \mathbf{B}^\top$ donde $\mathbf{B} \in \mathbb{R}^{n \times r}$ de rango r . Sea \mathbf{Y} vector aleatorio r -dimensional con $\mathbf{E}(\mathbf{Y}) = \mathbf{0}$ y $\text{Cov}(\mathbf{Y}) = \mathbf{I}$. Haciendo $\mathbf{X} = \mathbf{B} \mathbf{Y}$, sigue que $\mathbf{E}(\mathbf{X}) = \mathbf{0}$ y

$$\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{B} \mathbf{Y}) = \mathbf{B} \text{Cov}(\mathbf{Y}) \mathbf{B}^\top = \mathbf{B} \mathbf{B}^\top.$$

Es decir, corresponde a una matriz de covarianza. \square

RESULTADO 1.8. *Sea \mathbf{X} vector aleatorio n -dimensional y considere la transformación lineal $\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{b}$, donde \mathbf{A} es una matriz de constantes $m \times n$ y \mathbf{b} es vector de constantes $m \times 1$. Entonces*

$$\mathbf{E}(\mathbf{Y}) = \mathbf{A} \mathbf{E}(\mathbf{X}) + \mathbf{b}, \quad \text{Cov}(\mathbf{Y}) = \mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}^\top.$$

EJEMPLO 1.9. Sea \mathbf{X} vector aleatorio n -dimensional con media $\mathbf{E}(\mathbf{X}) = \boldsymbol{\mu}$ y matriz de dispersión $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$. Sea

$$\boldsymbol{\Sigma} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top$$

la descomposición espectral de $\boldsymbol{\Sigma}$, donde \mathbf{U} es matriz ortogonal y $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$, y considere la siguiente transformación

$$\mathbf{Z} = \boldsymbol{\Lambda}^{-1/2} \mathbf{U}^\top (\mathbf{X} - \boldsymbol{\mu})$$

de este modo, obtenemos que

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

En efecto, la transformación $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu})$ también satisface que $\mathbf{E}(\mathbf{Z}) = \mathbf{0}$ y $\text{Cov}(\mathbf{Z}) = \mathbf{I}$.

Suponga que \mathbf{Z} es una matriz aleatoria $n \times p$ cuyas filas son vectores aleatorios independientes $p \times 1$, cada uno con la misma matriz de covarianza $\mathbf{\Sigma}$. Considere la partición

$$\mathbf{Z}^\top = (\mathbf{Z}_1, \dots, \mathbf{Z}_n),$$

donde $\text{Cov}(\mathbf{Z}_i) = \mathbf{\Sigma}$, para $i = 1, \dots, n$. Tenemos que

$$\text{vec}(\mathbf{Z}^\top) = \begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_n \end{pmatrix},$$

y dado que todos los \mathbf{Z}_i son independientes con la misma matriz de covarianza sigue que

$$\text{Cov}(\text{vec}(\mathbf{Z}^\top)) = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma} \end{pmatrix} = \mathbf{I}_n \otimes \mathbf{\Sigma}.$$

Ahora suponga que llevamos a cabo la transformación lineal $\mathbf{Y} = \mathbf{AZB}$, donde $\mathbf{A} \in \mathbb{R}^{r \times n}$, $\mathbf{B} \in \mathbb{R}^{p \times q}$ son matrices de constantes. Entonces $\mathbf{E}(\mathbf{Y}) = \mathbf{A} \mathbf{E}(\mathbf{Z}) \mathbf{B}$, mientras que

$$\text{vec}(\mathbf{Y}^\top) = (\mathbf{A} \otimes \mathbf{B}^\top) \text{vec}(\mathbf{Z}^\top),$$

de modo que

$$\mathbf{E}(\text{vec}(\mathbf{Y}^\top)) = (\mathbf{A} \otimes \mathbf{B}^\top) \mathbf{E}(\text{vec}(\mathbf{Z}^\top)).$$

Además tenemos que

$$\begin{aligned} \text{Cov}(\text{vec}(\mathbf{Y}^\top)) &= (\mathbf{A} \otimes \mathbf{B}^\top) \text{Cov}(\text{vec}(\mathbf{Z}^\top)) (\mathbf{A} \otimes \mathbf{B}^\top)^\top \\ &= (\mathbf{A} \otimes \mathbf{B}^\top) (\mathbf{I}_n \otimes \mathbf{\Sigma}) (\mathbf{A}^\top \otimes \mathbf{B}) \\ &= \mathbf{A} \mathbf{A}^\top \otimes \mathbf{B}^\top \mathbf{\Sigma} \mathbf{B}. \end{aligned}$$

DEFINICIÓN 1.10 (Matriz de correlación). Sea $\mathbf{X} = (X_1, \dots, X_p)^\top$ vector aleatorio con media $\boldsymbol{\mu}$ y matriz de covarianza $\mathbf{\Sigma}$. Se define la matriz de correlaciones como $\mathbf{R} = (\rho_{ij})$, donde

$$\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\{\text{var}(X_i) \text{var}(X_j)\}^{1/2}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}, \quad i, j = 1, \dots, p.$$

Note que, para $\mathbf{\Sigma}$ matriz de covarianza del vector aleatorio \mathbf{X} y con $\mathbf{D} = \text{diag}(\mathbf{\Sigma}) (= \text{diag}(\sigma_{11}, \dots, \sigma_{pp}))$ podemos escribir

$$\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{\Sigma} \mathbf{D}^{-1/2}.$$

Cada elemento de la diagonal de \mathbf{R} es igual a 1, mientras que sus elementos fuera de la diagonal están entre -1 y 1. Además se desprende desde la definición que \mathbf{R} es una matriz semidefinida positiva.

RESULTADO 1.11. Sea \mathbf{X} vector aleatorio p -dimensional con $\mathbf{E}(\mathbf{X}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{X}) = \mathbf{\Sigma}$. Sea \mathbf{A} una matriz $p \times p$. Entonces

$$\mathbf{E}(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{A} \mathbf{\Sigma}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}.$$

DEMOSTRACIÓN. Tenemos

$$\begin{aligned}\mathbf{E}(\mathbf{X}^\top \mathbf{A} \mathbf{X}) &= \mathbf{E}(\text{tr } \mathbf{X}^\top \mathbf{A} \mathbf{X}) = \mathbf{E}(\text{tr } \mathbf{A} \mathbf{X} \mathbf{X}^\top) \\ &= \text{tr } \mathbf{E}(\mathbf{A} \mathbf{X} \mathbf{X}^\top) = \text{tr } \mathbf{A} \mathbf{E}(\mathbf{X} \mathbf{X}^\top) \\ &= \text{tr } \mathbf{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top) = \text{tr}(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}.\end{aligned}$$

□

Considere el siguiente caso especial: sea $\mathbf{Y} = \mathbf{X} - \mathbf{a}$, entonces $\text{Cov}(\mathbf{Y}) = \text{Cov}(\mathbf{X})$ y tenemos

$$\mathbf{E}[(\mathbf{X} - \mathbf{a})^\top \mathbf{A}(\mathbf{X} - \mathbf{a})] = \text{tr}(\mathbf{A} \boldsymbol{\Sigma}) + (\boldsymbol{\mu} - \mathbf{a})^\top \mathbf{A}(\boldsymbol{\mu} - \mathbf{a}).$$

EJEMPLO 1.12. Sea $\mathbf{1}_n = (1, \dots, 1)^\top$ vector n -dimensional cuyos componentes son todos 1. Note que, $\mathbf{1}_n^\top \mathbf{1}_n = n$. Considere el vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)^\top$, entonces

$$\mathbf{X}^\top \mathbf{X} = \sum_{i=1}^n X_i^2, \quad \mathbf{1}^\top \mathbf{X} = \sum_{i=1}^n X_i.$$

De este modo, tenemos

$$\begin{aligned}\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \mathbf{X}^\top \mathbf{X} - n\left(\frac{1}{n} \mathbf{1}^\top \mathbf{X}\right)^2 \\ &= \mathbf{X}^\top \mathbf{X} - n\left(\frac{1}{n} \mathbf{1}^\top \mathbf{X}\right)\left(\frac{1}{n} \mathbf{1}^\top \mathbf{X}\right) = \mathbf{X}^\top \mathbf{X} - \frac{1}{n} \mathbf{X}^\top \mathbf{1} \mathbf{1}^\top \mathbf{X} \\ &= \mathbf{X}^\top \left(\mathbf{I} - \frac{1}{n} \mathbf{J}_n\right) \mathbf{X}, \quad \mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\top\end{aligned}$$

Llamaremos a $\mathbf{C} = \mathbf{I} - \frac{1}{n} \mathbf{J}_n$ la *matriz de centrado*. Suponga que X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 , sigue que

$$\mathbf{E}(\mathbf{X}) = \mu \mathbf{1}_n, \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n,$$

pues $\text{Cov}(X_i, X_j) = 0$ ($i \neq j$). Por tanto, podemos usar el Resultado (1.11) para calcular la esperanza de la variable aleatoria,

$$Q = \sum_{i=1}^n (X_i - \bar{X})^2 = \mathbf{X}^\top \mathbf{C} \mathbf{X},$$

obteniendo

$$\mathbf{E}(Q) = \sigma^2 \text{tr}(\mathbf{C}) + \mu^2 \mathbf{1}^\top \mathbf{C} \mathbf{1}.$$

EJEMPLO 1.13. Sea X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 . Considere la forma cuadrática

$$Q = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2.$$

De este modo

$$Q = \mathbf{X}^\top \mathbf{C} \mathbf{X} = \mathbf{X}^\top \left(\mathbf{I} - \frac{1}{n} \mathbf{J}_n\right) \mathbf{X},$$

como $\text{Cov}(X_i, X_j) = 0$ ($i \neq j$), $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$ y $\mathbf{E}(\mathbf{X}) = \mu \mathbf{1}_n$, obtenemos que

$$\mathbf{E}(Q) = \sigma^2 \text{tr}(\mathbf{C}) + \mu^2 \mathbf{1}^\top \mathbf{C} \mathbf{1} = \sigma^2(n-1).$$

Es fácil verificar que

$$\begin{aligned}\text{tr}(\mathbf{C}) &= \text{tr}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right) = \text{tr}(\mathbf{I}) - \frac{1}{n}\text{tr}(\mathbf{1}\mathbf{1}^\top) = n - \frac{1}{n}\mathbf{1}^\top\mathbf{1} = n - 1, \\ \mathbf{C}\mathbf{1} &= \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right)\mathbf{1} = \mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{1} = \mathbf{1} - \mathbf{1} = \mathbf{0},\end{aligned}$$

de donde sigue que $\mathbb{E}(Q) = \sigma^2(n-1)$.

RESULTADO 1.14. *Si \mathbf{X} es vector aleatorio $n \times 1$. Entonces su distribución está determinada por las distribuciones de las funciones lineales $\mathbf{a}^\top \mathbf{X}$, para todo $\mathbf{a} \in \mathbb{R}^n$.*

DEMOSTRACIÓN. La función característica de $\mathbf{a}^\top \mathbf{X}$ es

$$\varphi_{\mathbf{a}^\top \mathbf{X}}(t) = \mathbb{E}\{\exp(it\mathbf{a}^\top \mathbf{X})\},$$

de modo que

$$\varphi_{\mathbf{a}^\top \mathbf{X}}(1) = \mathbb{E}\{\exp(i\mathbf{a}^\top \mathbf{X})\} \quad (= \varphi_X(\mathbf{a})).$$

Es considerada como una función de \mathbf{a} , esto es, la función característica (conjunta) de \mathbf{X} . El resultado sigue notando que una distribución en \mathbb{R}^n está completamente determinada por su función característica. \square

La función característica permite un método bastante operativo para el cálculo del k -ésimo momentos de un vector aleatorio \mathbf{X} . En efecto,

$$\begin{aligned}\boldsymbol{\mu}_k(\mathbf{X}) &= \begin{cases} \mathbb{E}(\mathbf{X} \otimes \mathbf{X}^\top \otimes \cdots \otimes \mathbf{X}^\top), & k \text{ par}, \\ \mathbb{E}(\mathbf{X} \otimes \mathbf{X}^\top \otimes \cdots \otimes \mathbf{X}^\top \otimes \mathbf{X}), & k \text{ impar}, \end{cases} \\ &= \begin{cases} i^{-k} \frac{\partial^k \varphi(\mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^\top \cdots \partial \mathbf{t}^\top} \Big|_{\mathbf{t}=\mathbf{0}}, & k \text{ par}, \\ i^{-k} \frac{\partial^k \varphi(\mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^\top \cdots \partial \mathbf{t}^\top \partial \mathbf{t}} \Big|_{\mathbf{t}=\mathbf{0}}, & k \text{ impar}. \end{cases}\end{aligned}$$

1.3. Independencia de vectores aleatorios

Sea $\mathbf{Z} = (\mathbf{X}^\top, \mathbf{Y}^\top)^\top$ con \mathbf{X}, \mathbf{Y} vectores aleatorios n y q -dimensionales, respectivamente. Se dicen independientes si y sólo si

$$F(\mathbf{x}, \mathbf{y}) = G(\mathbf{x})H(\mathbf{y}),$$

donde $F(\mathbf{z})$, $G(\mathbf{x})$ y $H(\mathbf{y})$ son las funciones de distribución de \mathbf{Z} , \mathbf{X} e \mathbf{Y} , respectivamente.

Si \mathbf{Z} , \mathbf{X} e \mathbf{Y} tienen densidades $f(\mathbf{z})$, $g(\mathbf{x})$ y $h(\mathbf{y})$, respectivamente. Entonces \mathbf{X} e \mathbf{Y} son independientes si

$$f(\mathbf{z}) = g(\mathbf{x})h(\mathbf{y}),$$

o bien,

$$f(\mathbf{x}|\mathbf{y}) = g(\mathbf{x}).$$

RESULTADO 1.15. *Sean \mathbf{X} e \mathbf{Y} dos vectores aleatorios independientes. Entonces para funciones cualquiera κ y τ , tenemos*

$$\mathbb{E}\{\kappa(\mathbf{X})\tau(\mathbf{Y})\} = \mathbb{E}\{\kappa(\mathbf{X})\}\mathbb{E}\{\tau(\mathbf{Y})\},$$

si las esperanzas existen.

DEMOSTRACIÓN. En efecto, es fácil notar que

$$\begin{aligned} \mathbb{E}\{\kappa(\mathbf{X})\tau(\mathbf{Y})\} &= \int \int \kappa(\mathbf{x})\tau(\mathbf{y})g(\mathbf{x})h(\mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \\ &= \left(\int \kappa(\mathbf{x})g(\mathbf{x}) \, d\mathbf{x} \right) \left(\int \tau(\mathbf{y})h(\mathbf{y}) \, d\mathbf{y} \right) \\ &= \mathbb{E}\{\kappa(\mathbf{X})\} \mathbb{E}\{\tau(\mathbf{Y})\}. \end{aligned}$$

□

1.4. Cambios de variable

Considere la función $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, el *Jacobiano* se define como el valor absoluto del determinante de $D\mathbf{f}(\mathbf{x})$ y es denotado por

$$J(\mathbf{y} \rightarrow \mathbf{x}) = |D\mathbf{f}(\mathbf{x})|_+ = \text{abs}(\det(D\mathbf{f}(\mathbf{x}))),$$

donde $\mathbf{y} = \mathbf{f}(\mathbf{x})$. Note que si $\mathbf{z} = \mathbf{f}(\mathbf{y})$ y $\mathbf{y} = \mathbf{g}(\mathbf{x})$, entonces tenemos

$$J(\mathbf{z} \rightarrow \mathbf{x}) = J(\mathbf{z} \rightarrow \mathbf{y}) \cdot J(\mathbf{y} \rightarrow \mathbf{x})$$

$$J(\mathbf{y} \rightarrow \mathbf{x}) = \{J(\mathbf{x} \rightarrow \mathbf{y})\}^{-1}$$

El siguiente resultado presenta una de aplicación del Jacobiano de una transformación para obtener la función de densidad de una transformación de un vector aleatorio.

PROPOSICIÓN 1.16 (Transformación de vectores aleatorios continuos). *Sea \mathbf{X} vector aleatorio n -dimensional con densidad $f_X(\mathbf{x})$ y soporte $S = \{\mathbf{x} : f_X(\mathbf{x}) > 0\}$. Para $\mathbf{g} : S \rightarrow \mathbb{R}^n$ diferenciable e invertible, sea $\mathbf{y} = \mathbf{g}(\mathbf{x})$. Entonces la densidad de \mathbf{Y} está dada por*

$$\begin{aligned} f_Y(\mathbf{y}) &= |D\mathbf{g}^{-1}(\mathbf{y})|_+ f_X(\mathbf{g}^{-1}(\mathbf{y})) \\ &= \{J(\mathbf{y} \rightarrow \mathbf{x})\}^{-1} f_X(\mathbf{g}^{-1}(\mathbf{y})). \end{aligned}$$

EJEMPLO 1.17. Sea $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B}$, $\mathbf{Y} \in \mathbb{R}^{n \times q}$, $\mathbf{X} \in \mathbb{R}^{n \times q}$, $\mathbf{A} \in \mathbb{R}^{n \times n}$ y $\mathbf{B} \in \mathbb{R}^{q \times q}$. Entonces

$$d\mathbf{Y} = \mathbf{A}(d\mathbf{X})\mathbf{B},$$

vectorizando obtenemos

$$\text{vec } d\mathbf{Y} = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec } d\mathbf{X},$$

esto es, $D\mathbf{F}(\mathbf{X}) = \mathbf{B}^\top \otimes \mathbf{A}$, por tanto

$$J(\mathbf{Y} \rightarrow \mathbf{X}) = |\mathbf{B}^\top \otimes \mathbf{A}|_+ = |\mathbf{A}|_+^q |\mathbf{B}^\top|_+^n = |\mathbf{A}|_+^q |\mathbf{B}|_+^n$$

1.5. Distribución normal multivariada

La distribución normal multivariada ocupa un rol central en inferencia multivariada así como en modelación estadística. En esta sección introducimos la distribución normal multivariada mediante tres definiciones equivalentes, definiciones que permiten el estudio de las propiedades fundamentales de la distribución normal multivariada.

Una variable aleatoria normal (uni-dimensional) Z tiene una distribución normal con media cero y varianza uno si su función de densidad es de la forma

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

en cuyo caso escribimos $Z \sim \mathbf{N}(0, 1)$. Más generalmente una variable aleatoria $Y \in \mathbb{R}$ tiene distribución normal con media $\mu \in \mathbb{R}$ y varianza $\sigma^2 \geq 0$ si

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

en cuyo caso escribimos $Y \sim \mathbf{N}(\mu, \sigma^2)$. Cuando $\sigma^2 = 0$, la distribución $\mathbf{N}(\mu, \sigma^2)$ se interpreta como una distribución degenerada en μ . Si $Y \sim \mathbf{N}(\mu, \sigma^2)$, entonces su función característica adopta la forma

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

Sea Z_1, \dots, Z_n variables aleatorias independientes cada una con distribución $\mathbf{N}(0, 1)$ y considere el vector aleatorio $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$. De este modo, la densidad conjunta de \mathbf{Z} es dada por

$$\begin{aligned} f(\mathbf{z}) &= \prod_{i=1}^n (2\pi)^{-1/2} \exp(-\frac{1}{2}z_i^2) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n z_i^2\right) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\|\mathbf{z}\|^2\right), \end{aligned}$$

y anotamos $\mathbf{Z} \sim \mathbf{N}_n(\mathbf{0}, \mathbf{I})$.

DEFINICIÓN 1.18. Un vector aleatorio p -dimensional, \mathbf{X} tiene distribución normal con vector de medias $\boldsymbol{\mu} \in \mathbb{R}^p$ y matriz de covarianza $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma} \geq \mathbf{0}$ sólo si, para todo vector \mathbf{t} la variable aleatoria (uni-dimensional) $\mathbf{t}^\top \mathbf{X}$ es normal univariada, en cuyo caso escribimos $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

OBSERVACIÓN 1.19. Note que en la definición anterior *no* se ha hecho supuestos respecto de la independencia de los componentes de \mathbf{X} .

RESULTADO 1.20. Suponga que $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere la transformación lineal $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ donde $\mathbf{A} \in \mathbb{R}^{m \times p}$ con $\text{rg}(\mathbf{A}) = m$. Entonces $\mathbf{Y} \sim \mathbf{N}_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.

DEMOSTRACIÓN. Sea $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ y simplemente 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 1.18 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. \square

A partir del resultado anterior sigue que todas las distribuciones marginales de \mathbf{X} también son normalmente distribuidas. En particular, también permite apreciar que la distribución normal satisface la siguiente propiedad relativa a la simetría multivariada:

$$\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \sigma^2 \mathbf{I}_p) \implies \mathbf{Q}\mathbf{Z} \stackrel{d}{=} \mathbf{Z}, \quad \forall \mathbf{Q} \in \mathcal{O}_p.$$

RESULTADO 1.21. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces la función característica de \mathbf{X} es dada por

$$\varphi_X(\mathbf{t}) = \exp(it^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}).$$

DEMOSTRACIÓN. Sabemos que la función característica de un vector aleatorio, satisface

$$\varphi_X(\mathbf{t}) = \mathbb{E}\{\exp(it^\top \mathbf{X})\} = \varphi_{t^\top \mathbf{X}}(1),$$

donde la función característica de la variable aleatoria uni-dimensional $Y = \mathbf{t}^\top \mathbf{X}$ es evaluada en 1. Como $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ sólo si $\mathbf{t}^\top \mathbf{X} \sim \mathbf{N}_1(\mathbf{t}^\top \boldsymbol{\mu}, \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t})$, tenemos

$$\varphi_X(\mathbf{t}) = \exp(it^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}).$$

En efecto, sea Σ 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

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

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 $\Sigma = \mathbf{B}\mathbf{B}^\top$ y $\boldsymbol{\mu} \in \mathbb{R}^p$. Entonces \mathbf{X} tiene función característica

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

□

OBSERVACIÓN 1.22. El Resultado 1.20 puede ser demostrado de manera bastante simple usando la función característica (ver Ejercicio 1.3).

RESULTADO 1.23. Si $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I})$. Entonces

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

DEMOSTRACIÓN. Para mostrar el resultado deseado, podemos calcular el primer y segundo diferencial de la función característica del vector aleatorio $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I})$. De este modo,

$$d\varphi_Z(\mathbf{t}) = -\varphi_Z(\mathbf{t})\mathbf{t}^\top d\mathbf{t},$$

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

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

Ahora, el vector de medias y matriz de covarianzas están dadas por

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

□

OBSERVACIÓN 1.24. Considere

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{B}\mathbf{Z}, \quad \boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^\top,$$

con $\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{I})$. Usando los Resultados 1.20 y 1.23, sigue que

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

RESULTADO 1.25. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces la distribución marginal de cualquier subconjunto de k ($< p$) componentes de \mathbf{X} es normal k -variada.

DEMOSTRACIÓN. Considere la siguiente partición:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (1.3)$$

donde \mathbf{X}_1 y $\boldsymbol{\mu}_1$ son vectores $k \times 1$ y $\boldsymbol{\Sigma}_{11}$ es $k \times k$. Aplicando el Resultado 1.20 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 $\mathbf{X}_1 \sim \mathbf{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$. \square

Una consecuencia de este resultado es que la distribución marginal de *cada* componente de \mathbf{X} es normal univariada.

OBSERVACIÓN 1.26. La inversa del Resultado 1.25 *no* es verdad en general. Es decir, que cada componente de un vector aleatorio tenga distribución normal no implica que todo el vector siga una distribución normal multivariada.

RESULTADO 1.27. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y \mathbf{X} , $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son particionadas como en la Ecuación (1.3). 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

$$\begin{aligned} \varphi_{\mathbf{X}}(\mathbf{t}) &= \exp(i\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}) \\ &= \exp(it_1^\top \boldsymbol{\mu}_1 + it_2^\top \boldsymbol{\mu}_2 - \frac{1}{2}\mathbf{t}_1^\top \boldsymbol{\Sigma}_{11} \mathbf{t}_1 - \frac{1}{2}\mathbf{t}_2^\top \boldsymbol{\Sigma}_{22} \mathbf{t}_2) \\ &= \exp(it_1^\top \boldsymbol{\mu}_1 - \frac{1}{2}\mathbf{t}_1^\top \boldsymbol{\Sigma}_{11} \mathbf{t}_1) \exp(it_2^\top \boldsymbol{\mu}_2 - \frac{1}{2}\mathbf{t}_2^\top \boldsymbol{\Sigma}_{22} \mathbf{t}_2) \\ &= \varphi_{\mathbf{X}_1}(\mathbf{t}_1) \varphi_{\mathbf{X}_2}(\mathbf{t}_2), \end{aligned}$$

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

DEFINICIÓN 1.28. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\boldsymbol{\Sigma}$ es definida positiva, entonces la densidad de \mathbf{X} asume la forma

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

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

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

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

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

y $d\mathbf{Z} = d\mathbf{g}^{-1}(\mathbf{X}) = \mathbf{B}^{-1} d\mathbf{X}$, 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 $\boldsymbol{\Sigma}^{-1} = \mathbf{B}^{-\top} \mathbf{B}^{-1}$ sigue el resultado deseado. \square

EJEMPLO 1.29. Sea $\mathbf{X} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$ donde

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

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

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

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

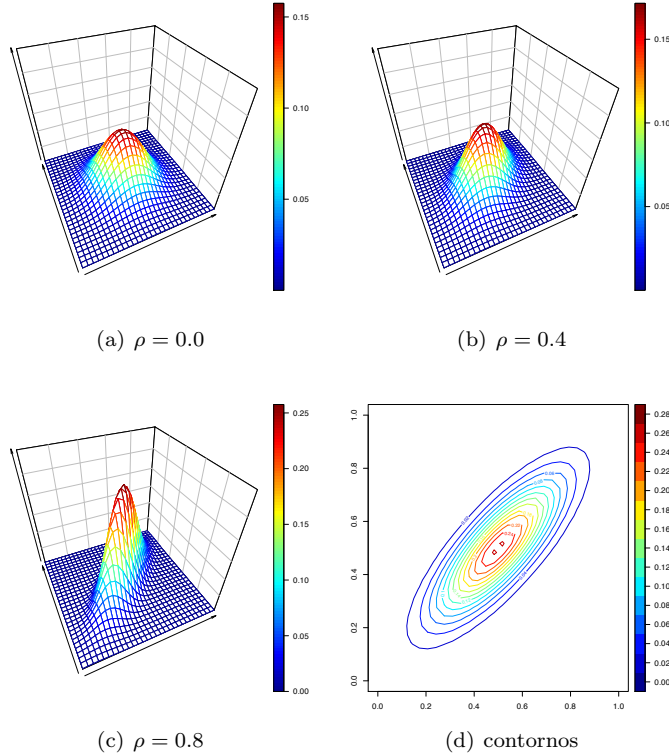


FIGURA 1. Densidad de $\mathbf{X} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$ para $\rho = 0.0, 0.4$ y 0.8 .

Es fácil apreciar 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

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}^\top \mathbf{Z} = \sum_{i=1}^p Z_i^2, \quad (1.4)$$

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* de \mathbf{X} a $\boldsymbol{\mu}$.

OBSERVACIÓN 1.30. Para la existencia de densidad hemos asumido que $\boldsymbol{\Sigma} > \mathbf{0}$. En el caso de que $\boldsymbol{\Sigma} \geq \mathbf{0}$ decimos que \mathbf{X} sigue una distribución normal singular.

Para introducir una definición de la función de densidad asociada a una variable con distribución normal singular, note que $X \sim \mathcal{N}(\mu, \sigma^2)$ con $\sigma^2 = 0 \Leftrightarrow x = \mu$ con probabilidad 1 (pues si $\sigma^2 = 0$, $P(X = \mu) = \lim_{n \rightarrow \infty} P(|X - \mu| < 1/n) = 0$, $\forall n$).

Considere $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\text{rg}(\boldsymbol{\Sigma}) = r < p$. Entonces, podemos escribir

$$\boldsymbol{\Sigma} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top = (\mathbf{U}_1, \mathbf{U}_2) \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} = \mathbf{U}_1 \boldsymbol{\Lambda}_1 \mathbf{U}_1^\top,$$

donde $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_r)$. De este modo, es claro que

$$\mathbf{U}^\top \boldsymbol{\Sigma} \mathbf{U} \implies \mathbf{U}_2^\top \boldsymbol{\Sigma} \mathbf{U}_2 = \mathbf{0},$$

es decir, tenemos que $\mathbf{U}_2^\top (\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0}$ con probabilidad 1. Mientras que

$$\mathbf{U}_1^\top (\mathbf{Y} - \boldsymbol{\mu}) \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Lambda}_1).$$

Además $\boldsymbol{\Sigma}^- = \mathbf{U}_1 \boldsymbol{\Lambda}_1^{-1} \mathbf{U}_1^\top = \mathbf{U}_1 (\mathbf{U}_1^\top \boldsymbol{\Sigma} \mathbf{U}_1)^{-1} \mathbf{U}_1^\top$. Así, \mathbf{Y} tiene la siguiente densidad normal (singular)

$$\begin{aligned} f_Y(\mathbf{y}) &= |2\pi \mathbf{U}_1^\top \boldsymbol{\Sigma} \mathbf{U}_1|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{U}_1^\top (\mathbf{y} - \boldsymbol{\mu}))^\top (\mathbf{U}_1^\top \boldsymbol{\Sigma} \mathbf{U}_1)^{-1} \mathbf{U}_1^\top (\mathbf{y} - \boldsymbol{\mu})\right\} \\ &= (2\pi)^{-r/2} |\boldsymbol{\Lambda}_1|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{U}_1 \boldsymbol{\Lambda}_1^{-1} \mathbf{U}_1^\top (\mathbf{y} - \boldsymbol{\mu})\right\}. \end{aligned}$$

El siguiente resultado presenta la distribución condicional de un vector aleatorio con distribución normal multivariada.

RESULTADO 1.31. Sea $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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 $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^- \boldsymbol{\Sigma}_{21}$. Entonces

- (a) $\mathbf{X}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^- \mathbf{X}_2 \sim \mathcal{N}_k(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^- \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11.2})$ y es independiente de \mathbf{X}_2 .
- (b) La distribución condicional

$$(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) \sim \mathcal{N}_k(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^- (\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11.2}).$$

DEMOSTRACIÓN. Considere la transformación lineal

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

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

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

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

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

sigue que $\boldsymbol{\Sigma}_{12}\mathbf{B}^\top = \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{B}^\top$ (y análogamente $\mathbf{B}\boldsymbol{\Sigma}_{21} = \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{B}^\top$). Esto es, si \mathbf{B} es escogida como $\mathbf{B} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-$, entonces \mathbf{Y}_1 y \mathbf{Y}_2 son independientes con distribución conjunta

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^- \mathbf{X}_2 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^- \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11.2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right).$$

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

$$\begin{aligned} g(\mathbf{y}_1; \boldsymbol{\delta}_{1.2}, \boldsymbol{\Sigma}_{11.2}) &= |2\pi\boldsymbol{\Sigma}_{11.2}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_1 - \boldsymbol{\delta}_{1.2})^\top \boldsymbol{\Sigma}_{11.2}^{-1}(\mathbf{y}_1 - \boldsymbol{\delta}_{1.2})\right\} \\ f_2(\mathbf{y}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) &= |2\pi\boldsymbol{\Sigma}_{22}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2)\right\}, \end{aligned}$$

y la densidad conjunta para $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ adopta la forma

$$f(\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = g(\mathbf{y}_1; \boldsymbol{\delta}_{1.2}, \boldsymbol{\Sigma}_{11.2}) f_2(\mathbf{y}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}).$$

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 $g(\mathbf{y}_1; \boldsymbol{\delta}_{1.2}, \boldsymbol{\Sigma}_{11.2})$. Además, es fácil notar que la forma cuadrática

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

donde

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

lo que muestra el resultado. \square

OBSERVACIÓN 1.32. La esperanza de la distribución condicional de \mathbf{X}_1 dado \mathbf{X}_2 , es decir

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

se denomina *función de regresión* de \mathbf{X}_1 sobre \mathbf{X}_2 con coeficientes de regresión $\mathbf{B} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-$. Esta es una función lineal de \mathbf{X}_2 y la matriz de covarianza $\boldsymbol{\Sigma}_{11.2}$ no depende de \mathbf{X}_2 .

RESULTADO 1.33. Sea $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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

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

Este resultado permite obtener una condición para la independencia entre dos formas lineales en variables aleatorias normales, estos es \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 1.34. Considere X_1, \dots, X_n una muestra aleatoria desde $\mathbf{N}(\mu, \sigma^2)$ 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 $\mathbf{X} \sim \mathbf{N}_n(\mu \mathbf{1}, \sigma^2 \mathbf{I}_n)$ y \bar{X} con \mathbf{Z} son independientes pues $\mathbf{C} \mathbf{1} = \mathbf{0}$.

EJEMPLO 1.35. Sea $\mathbf{X} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$ y considere las transformaciones $\mathbf{Y}_1 = \mathbf{A} \mathbf{X}$ y $\mathbf{Y}_2 = (\mathbf{I} - \mathbf{A}^+ \mathbf{A})^\top \mathbf{X}$. 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.

1.6. Alternativas a la distribución normal multivariada

La distribución normal multivariada es de importancia fundamental en la teoría clásica de modelos lineales así como para análisis multivariado. A pesar de su uso amplio, es bien sabido que la inferencia estadística basada en la distribución normal es vulnerable a la presencia de datos atípicos, esto ha motivado considerar distribuciones alternativas que eviten este tipo de limitaciones. En esta dirección, varios autores han sugerido la clase de distribuciones elípticas (ver, por ejemplo [Fang et al., 1990](#); [Arellano, 1994](#)) particularmente debido al hecho de incluir distribuciones con colas más pesadas que la normal, tales como la t de Student, exponencial potencia y normal contaminada, entre otras. Una subclase importante de la familia de distribuciones elípticas es la clase de distribuciones de mezcla de escala normal ([Andrews y Mallows, 1974](#)) la que tiene propiedades similares a la distribución normal, es relativamente simple de trabajar y permite proponer procedimientos para estimación robusta. A continuación se presenta la definición y algunos ejemplos de distribuciones en la clase elíptica.

DEFINICIÓN 1.36. Sea \mathbf{U} vector aleatorio $p \times 1$ con *distribución uniforme* sobre el conjunto

$$\mathcal{S}_p = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\| = 1\}, \quad (1.5)$$

esto es \mathcal{S}_p denota la *superficie de la esfera unitaria* en \mathbb{R}^p . En cuyo caso anotamos $\mathbf{U} \sim \mathbf{U}(\mathcal{S}_p)$.

PROPIEDAD 1.37. Si $\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{I})$, entonces $\mathbf{U} = (U_1, \dots, U_p)^\top \sim \mathbf{U}(\mathcal{S}_p)$, donde

$$\mathbf{U} = \frac{\mathbf{Z}}{\|\mathbf{Z}\|}.$$

El resultado anterior es muy relevante pues permite definir la densidad de un vector aleatorio $\mathbf{U} \sim \mathcal{U}(\mathcal{S}_p)$. Así como, por ejemplo, generar observaciones sobre la esfera unitaria. Considere el siguiente gráfico,

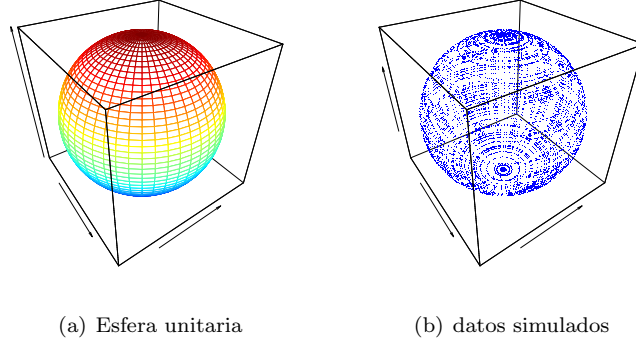


FIGURA 2. Esfera unitaria y datos generados sobre la superficie \mathcal{S}_p .

DEFINICIÓN 1.38. Un vector aleatorio $p \times 1$, \mathbf{X} se dice que tiene simetría esférica si para cualquier $\mathbf{Q} \in \mathcal{O}_p$, sigue que

$$\mathbf{Q}\mathbf{X} \stackrel{d}{=} \mathbf{X}.$$

EJEMPLO 1.39. Sea $\mathbf{U} \sim \mathcal{U}(\mathcal{S}_p)$, entonces es bastante obvio que $\mathbf{Q}\mathbf{U} \stackrel{d}{=} \mathbf{U}$.

EJEMPLO 1.40. Suponga $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I})$. Tenemos que

$$\mathbf{Q}\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I}),$$

para $\mathbf{Q} \in \mathcal{O}_p$, es decir $\mathbf{Q}\mathbf{X} \stackrel{d}{=} \mathbf{X}$ tiene simetría esférica.

DEFINICIÓN 1.41. Un vector aleatorio p -dimensional tiene *distribución esférica* sólo si su función característica satisface

- (a) $\varphi(\mathbf{Q}^\top \mathbf{t}) = \varphi(\mathbf{t})$, para todo $\mathbf{Q} \in \mathcal{O}_p$.
- (b) Existe una función $\psi(\cdot)$ de una variable escalar tal que $\varphi(\mathbf{t}) = \psi(\mathbf{t}^\top \mathbf{t})$.

En este caso escribimos $\mathbf{X} \sim \mathcal{S}_p(\psi)$.

EJEMPLO 1.42. Sea $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$, tenemos que

$$\varphi(\mathbf{t}) = \exp\{-\frac{1}{2}(t_1^2 + \cdots + t_p^2)\} = \exp(-\frac{1}{2}\mathbf{t}^\top \mathbf{t}).$$

RESULTADO 1.43. Sea $\psi(\mathbf{t}^\top \mathbf{t})$ la función característica del vector aleatorio \mathbf{X} . Entonces \mathbf{X} tiene representación estocástica

$$\mathbf{X} \stackrel{d}{=} \mathbf{R}\mathbf{U},$$

donde $\mathbf{U} \sim \mathcal{U}(\mathcal{S}_p)$ y $\mathbf{R} \sim F(\mathbf{X})$ son independientes.

RESULTADO 1.44. Suponga que $\mathbf{X} \stackrel{d}{=} \mathbf{R}\mathbf{U} \sim \mathcal{S}_p(\psi)$ ($P(\mathbf{X} = \mathbf{0}) = 0$), entonces

$$\|\mathbf{X}\| \stackrel{d}{=} \mathbf{R}, \quad \frac{\mathbf{X}}{\|\mathbf{X}\|} \stackrel{d}{=} \mathbf{U}.$$

Además $\|\mathbf{X}\|$ y $\mathbf{X}/\|\mathbf{X}\|$ son independientes.

RESULTADO 1.45. *El vector de medias y la matriz de covarianza de $\mathbf{U} \sim \mathbf{U}(\mathcal{S}_p)$ son:*

$$\mathbf{E}(\mathbf{U}) = \mathbf{0}, \quad \text{Cov}(\mathbf{U}) = \frac{1}{p} \mathbf{I}_p,$$

respectivamente.

DEMOSTRACIÓN. Sea $\mathbf{X} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{I})$, tenemos que $\mathbf{X} \stackrel{d}{=} \|\mathbf{X}\| \mathbf{U}$, con $\|\mathbf{X}\|$ independiente de \mathbf{U} . Sabemos que $\|\mathbf{X}\|^2 \sim \chi^2(p)$. Dado que

$$\mathbf{E}(\mathbf{X}) = \mathbf{0}, \quad \mathbf{E}(\|\mathbf{X}\|) > 0, \quad \text{y} \quad \mathbf{E}(\|\mathbf{X}\|^2) = p, \quad \text{Cov}(\mathbf{X}) = \mathbf{I}_p,$$

el resultado sigue. \square

RESULTADO 1.46. *Si $\mathbf{X} \stackrel{d}{=} R\mathbf{U} \sim \mathbf{S}_p(g)$ y $\mathbf{E}(R^2) < \infty$. Entonces,*

$$\mathbf{E}(\mathbf{X}) = \mathbf{0}, \quad \text{Cov}(\mathbf{X}) = \frac{\mathbf{E}(R^2)}{p} \mathbf{I}_p,$$

respectivamente.

DEMOSTRACIÓN. En efecto, como R y \mathbf{U} son independientes, sigue que

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}(R) \mathbf{E}(\mathbf{U}) = \mathbf{0},$$

$$\text{Cov}(\mathbf{X}) = \mathbf{E}(R^2) \mathbf{E}(\mathbf{U}\mathbf{U}^\top) = \mathbf{E}(R^2) \text{Cov}(\mathbf{U}) = \frac{\mathbf{E}(R^2)}{p} \mathbf{I}_p,$$

siempre que $\mathbf{E}(R) < \infty$ y $\mathbf{E}(R^2) < \infty$. \square

DEFINICIÓN 1.47. Un vector aleatorio $p \times 1$, \mathbf{X} tiene *distribución de contornos elípticos* con parámetros $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} \geq \mathbf{0}$ si

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{B}\mathbf{Y}, \quad \mathbf{Y} \sim \mathbf{S}_k(\psi),$$

donde $\mathbf{B} \in \mathbb{R}^{k \times p}$ es matriz de rango completo tal que, $\mathbf{B}\mathbf{B}^\top = \boldsymbol{\Sigma}$ con $\text{rg}(\boldsymbol{\Sigma}) = k$ y escribimos $\mathbf{X} \sim \text{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \psi)$.

OBSERVACIÓN 1.48. La función característica de $\mathbf{X} \sim \text{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \psi)$ es de la forma

$$\varphi(\mathbf{t}) = \exp(i\mathbf{t}^\top \boldsymbol{\mu}) \psi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}).$$

Note además que la representación estocástica de \mathbf{X} es dada por

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{B}\mathbf{U},$$

donde $R \geq 0$ es independiente de \mathbf{U} y $\mathbf{B}\mathbf{B}^\top = \boldsymbol{\Sigma}$.

RESULTADO 1.49. *Suponga que $\mathbf{X} \sim \text{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \psi)$ y $\mathbf{E}(R^2) < \infty$. Entonces*

$$\mathbf{E}(\mathbf{X}) = \boldsymbol{\mu}, \quad \text{Cov}(\mathbf{X}) = \frac{\mathbf{E}(R^2)}{p} \boldsymbol{\Sigma}.$$

DEFINICIÓN 1.50. Se dice que el vector \mathbf{X} tiene distribución de contornos elípticos si su función de densidad es de la forma

$$f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})), \quad \mathbf{x} \in \mathbb{R}^p,$$

donde $g : \mathbb{R} \rightarrow [0, \infty)$ es función decreciente, llamada *función generadora de densidad*, tal que:

$$\int_0^\infty u^{p/2-1} g(u) \, du < \infty,$$

y escribimos $\mathbf{X} \sim \text{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$.

OBSERVACIÓN 1.51. Asuma que $\mathbf{X} \sim \text{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \psi)$ con $\text{rg}(\boldsymbol{\Sigma}) = k$. Entonces,

$$U = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^- (\mathbf{X} - \boldsymbol{\mu}) \stackrel{d}{=} R^2,$$

donde $\boldsymbol{\Sigma}^-$ es una inversa generalizada de $\boldsymbol{\Sigma}$.

EJEMPLO 1.52. En la siguiente figura se presenta la densidad asociadas a las siguientes funciones g :

- Normal: $g(u) = c_1 \exp(-u/2)$.
- Laplace: $g(u) = c_2 \exp(-\sqrt{u}/2)$.
- Cauchy: $g(u) = c_3(1 + u)^{-(p+1)/2}$.
- Exponencial potencia (PE): $g(u) = c_4 \exp(-u^\lambda/2)$, $\lambda = 2$.

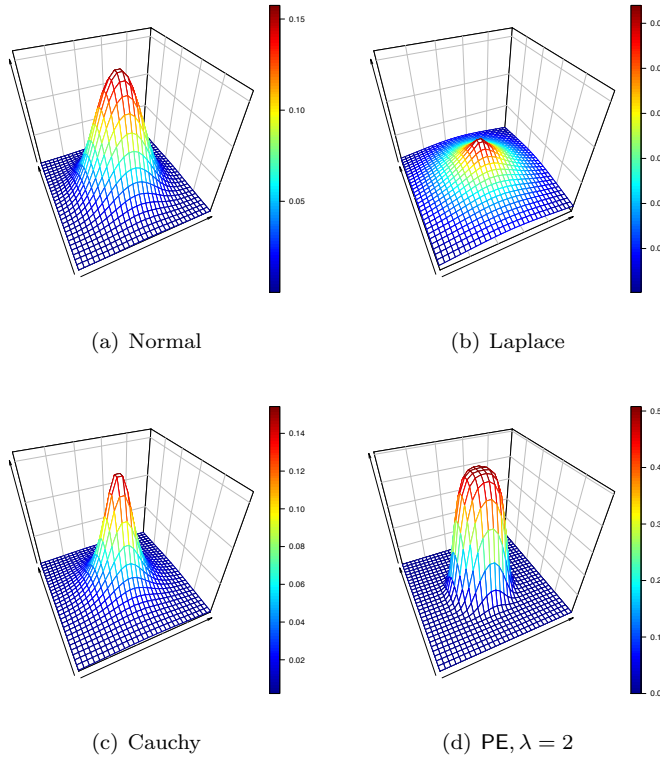


FIGURA 3. Funciones de densidad del vector $\mathbf{X} \sim \text{EC}_2(\mathbf{0}, \mathbf{I}; g)$ para las distribuciones normal, Laplace, Cauchy y exponencial potencia con $\lambda = 2$.

EJEMPLO 1.53 (Distribución t de Student). La función generadora de densidad de un vector aleatorio con distribución t de Student asume la forma

$$g(u) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{p/2}} \left(1 + \frac{u}{\nu}\right)^{-(\nu+p)/2}, \quad \nu > 0.$$

Para la distribución t de Student, tenemos que $R^2/p \sim F_{p,\nu}$. Además, la función característica de $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ es dada por

$$\varphi(\mathbf{t}) = \frac{\|\sqrt{\nu}\boldsymbol{\Sigma}^{1/2}\mathbf{t}\|^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)} \exp\{i\mathbf{t}^\top \boldsymbol{\mu}\} K_{\nu/2}(\|\sqrt{\nu}\boldsymbol{\Sigma}^{1/2}\mathbf{t}\|), \quad \mathbf{t} \in \mathbb{R}^p,$$

donde $K_\nu(x)$ denota la función de Bessel modificada de segundo tipo. Un caso particular importante corresponde a la distribución Cauchy, cuando $\nu = 1$, mientras que la distribución normal corresponde al caso límite $\nu \rightarrow \infty$.

EJEMPLO 1.54 (Distribución Exponencial Potencia). Para la distribución Exponencial Potencia (Gómez et al., 1988), la función generadora de densidades es dada por

$$g(u) = \frac{p\Gamma(\frac{p}{2})\pi^{-p/2}}{\Gamma(1 + \frac{p}{2\lambda})2^{1+\frac{p}{2\lambda}}} \exp(-u^\lambda/2), \quad \lambda > 0.$$

y es usual utilizar la notación $\mathbf{X} \sim \text{PE}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$. En este caso tenemos que la variable aleatoria positiva R tiene densidad

$$h(r) = \frac{p}{\Gamma(1 + \frac{p}{2\lambda})2^{\frac{p}{2\lambda}}} r^{p-1} \exp(-r^{2\lambda}/2), \quad r > 0.$$

Note también que $R^{2\lambda} \sim \text{Gama}(\frac{1}{2}, \frac{p}{2\lambda})$. Debemos destacar que esta clase de distribuciones contiene la distribución normal como un caso particular cuando $\lambda = 1$. Mientras que tiene colas más pesadas que la normal si $\lambda < 1$ y colas más livianas para el caso $\lambda > 1$. Otro caso particular de interés es la distribución Laplace, que es recuperada cuando $\lambda = 1/2$.

DEFINICIÓN 1.55. Sea $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma}$ matriz $p \times p$ definida positiva y \mathbf{H} función de distribución de la variable aleatoria positiva W . Entonces, se dice que el vector aleatorio \mathbf{X} sigue una *distribución de mezcla de escala normal* si su función de densidad asume la forma

$$f(\mathbf{x}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \int_0^\infty w^{p/2} \exp(-wu/2) d\mathbf{H}(w),$$

donde $u = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ y anotamos $\mathbf{X} \sim \text{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{H})$.

EJEMPLO 1.56 (Distribución Slash). Un vector aleatorio \mathbf{X} tiene distribución Slash si su función de densidad es de la forma:

$$f(\mathbf{x}) = \nu(2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \int_0^1 w^{p/2+\nu-1} \exp(-wu/2) dw.$$

En este caso, tenemos que $h(w) = \nu w^{\nu-1}$, para $w \in (0, 1)$ y $\nu > 0$. Es decir $W \sim \text{Beta}(\nu, 1)$.

OBSERVACIÓN 1.57. Un vector aleatorio $\mathbf{X} \sim \text{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{H})$ admite la representación

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + W^{-1/2} \mathbf{Z},$$

donde $\mathbf{Z} \sim \text{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ y $W \sim \mathbf{H}(\nu)$ son independientes. También podemos utilizar la siguiente estructura jerárquica:

$$\mathbf{X}|W \sim \text{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/W), \quad W \sim \mathbf{H}(\nu).$$

Esta representación permite, por ejemplo

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}(\mathbf{E}(\mathbf{X}|W)) = \boldsymbol{\mu}$$

$$\text{Cov}(\mathbf{X}) = \mathbf{E}(\text{Cov}(\mathbf{X}|W)) + \text{Cov}(\mathbf{E}(\mathbf{X}|W)) = \mathbf{E}(W^{-1})\boldsymbol{\Sigma},$$

siempre que $\mathbf{E}(W^{-1}) < \infty$.

EJEMPLO 1.58 (Distribución t de Student). Para $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, con $\nu > 0$, podemos escribir

$$\mathbf{X}|W \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/W), \quad W \sim \text{Gamma}(\nu/2, \nu/2),$$

es decir, la función de densidad asociado a la variable de mezcla, es dada por

$$h(\omega; \nu) = \frac{(\nu/2)^{\nu/2} \omega^{\nu/2-1}}{\Gamma(\nu/2)} \exp(-\nu\omega/2).$$

EJEMPLO 1.59 (Distribución normal contaminada). Considere $\mathbf{X} \sim \text{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \epsilon, \gamma)$ [Little \(1988\)](#) donde $0 \leq \epsilon \leq 1$ denota el *porcentaje de contaminación* y $0 < \gamma < 1$ corresponde a un *factor de inflación de escala*. En este caso, la variable de mezcla tiene densidad

$$h(\omega; \boldsymbol{\delta}) = \begin{cases} \epsilon, & \omega = \gamma \\ 1 - \epsilon & \omega = 1 \end{cases},$$

con $\boldsymbol{\delta} = (\epsilon, \gamma)^\top$. Podemos notar que la función de densidad adopta la forma:

$$f(\mathbf{x}) = (1 - \epsilon)|2\pi\boldsymbol{\Sigma}|^{-1/2} \exp(-u/2) + \epsilon\gamma^{p/2}|2\pi\boldsymbol{\Sigma}|^{-1/2} \exp(-\lambda u/2).$$

1.7. Algunas distribuciones no centrales

Las distribuciones chi-cuadrado, F , t de Student no central son derivadas desde la distribución normal multivariada y son útiles para desarrollar la inferencia en modelo de regresión lineal.

RESULTADO 1.60. Sea $\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{I})$ y sea $U = \mathbf{Z}^\top \mathbf{Z}$. Entonces $U \sim \chi^2(p)$, con función de densidad

$$f(u) = \frac{1}{2^{p/2}\Gamma(p/2)} u^{p/2-1} \exp(-u/2), \quad u > 0.$$

DEMOSTRACIÓN. Como U es una función de variables aleatorias normales, entonces su función característica asume la forma

$$\begin{aligned} \varphi_U(t) &= \mathbf{E}\{\exp(itU)\} = \int_{\mathbb{R}^p} \exp(it\mathbf{z}^\top \mathbf{z}) (2\pi)^{-p/2} \exp(-\tfrac{1}{2}\mathbf{z}^\top \mathbf{z}) d\mathbf{z} \\ &= (2\pi)^{-p/2} \int_{\mathbb{R}^p} \exp(-\tfrac{1}{2}(1 - 2it)\mathbf{z}^\top \mathbf{z}) d\mathbf{z} = (1 - 2it)^{-p/2}, \end{aligned}$$

que corresponde a la función característica de una variable aleatoria chi-cuadrado con p grados de libertad. \square

DEFINICIÓN 1.61 (Distribución chi-cuadrado no central). Si $\mathbf{Y} \sim \mathbf{N}_p(\boldsymbol{\mu}, \mathbf{I})$, entonces $U = \mathbf{Y}^\top \mathbf{Y}$ tiene *distribución chi-cuadrado no central* con p grados de libertad y parámetro de no centralidad $\lambda = \boldsymbol{\mu}^\top \boldsymbol{\mu}/2$, en cuyo caso anotamos $U \sim \chi^2(p; \lambda)$.

RESULTADO 1.62. Sea $\mathbf{Y} \sim \mathbf{N}_p(\boldsymbol{\mu}, \mathbf{I})$ donde $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p) \neq \mathbf{0}$ y sea $U = \mathbf{Y}^\top \mathbf{Y}$. Entonces la función característica de U es dada por

$$\varphi_U(t) = (1 - 2it)^{-p/2} \exp\left(\frac{2it\lambda}{1 - 2it}\right),$$

con $\lambda = \boldsymbol{\mu}^\top \boldsymbol{\mu}/2$.

DEMOSTRACIÓN. Como Y_1, \dots, Y_p son variables aleatorias independientes, tenemos

$$\begin{aligned} \varphi_U(t) &= \mathbb{E} \left\{ \exp \left(t \sum_{j=1}^n Y_j^2 \right) \right\} = \mathbb{E} \left\{ \prod_{j=1}^p \exp(tY_j^2) \right\} = \prod_{j=1}^p \mathbb{E} \{ \exp(tY_j^2) \} \\ &= \prod_{j=1}^p \varphi_{Y_j^2}(t). \end{aligned}$$

Ahora, la función característica asociada a la variable aleatoria Y_j^2 es dada por

$$\begin{aligned} \varphi_{Y_j^2}(t) &= \int_{-\infty}^{\infty} \exp(it y_j^2) (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(y_j - \mu_j)^2\right\} dy_j \\ &= \exp\left\{\frac{\mu_j^2}{2} \left(\frac{1}{1-2it}\right) - \frac{\mu_j^2}{2}\right\} \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\left\{-\frac{(1-2it)}{2} \left(y_j - \frac{\mu_j}{1-2it}\right)^2\right\} dy_j, \end{aligned}$$

de este modo,

$$\varphi_{Y_j^2}(t) = (1 - 2it)^{-1/2} \exp\left\{\frac{\mu_j^2}{2} \left(\frac{2it}{1 - 2it}\right)\right\},$$

y por tanto la función característica de la variable $U = \sum_{j=1}^p Y_j^2$, asume la forma

$$\varphi_U(t) = (1 - 2it)^{-p/2} \exp\left(\frac{2it\lambda}{1 - 2it}\right), \quad \lambda = \boldsymbol{\mu}^\top \boldsymbol{\mu}/2.$$

□

OBSERVACIÓN 1.63. Es interesante notar que la función característica de la variable $U = \mathbf{Y}^\top \mathbf{Y}$, puede ser escrita como

$$\begin{aligned} \varphi_U(t) &= (1 - 2it)^{-p/2} \exp\left(\frac{\lambda}{1 - 2it} - \lambda\right) \\ &= (1 - 2it)^{-p/2} e^{-\lambda} \sum_{k=0}^{\infty} \frac{\{\lambda/(1 - 2it)\}^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} (1 - 2it)^{-(p+2k)/2}. \end{aligned}$$

Es decir, la función característica de U es un *promedio ponderado con pesos Poisson* de funciones características de variables aleatorias chi-cuadrado con $p + 2k$ grados de libertad.

Usando la relación entre funciones características y sus correspondientes funciones de densidad, sigue que la chi-cuadrado no central tiene la siguiente representación de mezcla

$$U|Z \sim \chi^2(p + 2z), \quad Z \sim \text{Poisson}(\lambda), \quad (1.6)$$

con densidad

$$f(u) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \frac{1}{2^{p/2+k} \Gamma(\frac{p}{2} + k)} u^{p/2+k-1} \exp(-u/2), \quad u > 0.$$

La representación en (1.6) es muy útil para obtener los momentos de una variable aleatoria con distribución chi-cuadrado no central. En efecto, el valor esperado de $U \sim \chi^2(p; \lambda)$ es dado por

$$\mathbb{E}(U) = \mathbb{E}\{\mathbb{E}(U|Z)\} = \mathbb{E}\{p + 2Z\} = p + 2\mathbb{E}(Z) = p + 2\lambda,$$

mientras que la varianza de U puede ser calculada como

$$\begin{aligned} \text{var}(U) &= \mathbb{E}\{\text{var}(U|Z)\} + \text{var}\{\mathbb{E}(U|Z)\} \\ &= \mathbb{E}\{2(p + 2Z)\} + \text{var}(p + 2Z) \\ &= 2p + 4\lambda + 4\lambda = 2p + 8\lambda. \end{aligned}$$

RESULTADO 1.64. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ donde $\boldsymbol{\Sigma}$ es matriz no singular. Entonces

- (a) $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$.
- (b) $\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} \sim \chi^2(p; \lambda)$, donde $\lambda = \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$.

DEMOSTRACIÓN. La idea de la demostración se basa en transformar los componentes de \mathbf{X} en variables aleatorias normales independientes. Considere $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^\top$ con \mathbf{B} no singular. Para probar (a), tome

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

luego $\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{I})$ y de este modo

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}^\top \mathbf{Z} \sim \chi^2(p; 0).$$

Para probar (b), sea $\mathbf{Y} = \mathbf{B}^{-1} \mathbf{X}$, luego

$$\mathbf{Y} \sim \mathbf{N}_p(\mathbf{B}^{-1} \boldsymbol{\mu}, \mathbf{I}),$$

y

$$\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} = \mathbf{Y}^\top \mathbf{B}^\top \boldsymbol{\Sigma}^{-1} \mathbf{B} \mathbf{Y} = \mathbf{Y}^\top \mathbf{Y},$$

que por definición tiene una distribución chi-cuadrado no central, con parámetro de no centralidad

$$\lambda = \frac{1}{2} (\mathbf{B}^{-1} \boldsymbol{\mu})^\top (\mathbf{B}^{-1} \boldsymbol{\mu}) = \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}.$$

□

DEFINICIÓN 1.65 (Distribución F no central). Sea $X_1 \sim \chi^2(\nu_1; \lambda)$ y $X_2 \sim \chi^2(\nu_2)$ variables aleatorias independientes. Entonces,

$$F = \frac{X_1/\nu_1}{X_2/\nu_2} \sim F(\nu_1, \nu_2, \lambda),$$

es decir F sigue una *distribución F no central* con ν_1 y ν_2 grados de libertad y parámetro de no centralidad λ .

DEFINICIÓN 1.66 (Distribución Beta no central). Considere $U_1 \sim \chi^2(\nu_1, \lambda)$, $U_2 \sim \chi^2(\nu_2)$ tal que U_1 y U_2 son variables aleatorias independientes. Entonces,

$$G = \frac{U_1}{U_1 + U_2} \sim \text{Beta}(\nu_1, \nu_2, \lambda),$$

esto es, G sigue una *distribución Beta no central* con parámetros de forma y escala ν_1 y ν_2 , respectivamente y parámetro de no centralidad λ .

DEFINICIÓN 1.67 (Distribución t de Student no central). Si $Y \sim \mathbf{N}(\mu, \sigma^2)$ y $U/\sigma^2 \sim \chi^2(\nu)$ son independientes, entonces

$$T = \frac{Y}{\sqrt{U/\nu}} \sim t_\nu(\lambda), \quad \lambda = \mu/\sigma,$$

es llamada una variable aleatoria con *distribución t de Student no central* con ν grados de libertad y parámetro de no centralidad λ .

Note también que si $Z \sim \mathbf{N}(0, 1)$, $U \sim \chi^2(\nu)$, δ es una constante, y Z es independiente de U , entonces

$$T = \frac{Z + \delta}{\sqrt{U/\nu}} \sim t_\nu(\delta).$$

Además el cuadrado de una variable aleatoria t no central se distribuye como una variable aleatoria F no central con parámetro de no centralidad $\delta = \lambda^2/2$. De este modo,

$$t_\nu^2(\lambda) \stackrel{d}{=} F(1, \nu, \lambda^2/2).$$

1.8. Distribución de formas cuadráticas

Para motivar ideas, sabemos que si $\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{I})$, entonces $U = \mathbf{Z}^\top \mathbf{Z} \sim \chi^2(p)$ pues corresponde a la suma de variables aleatorias IID $\mathbf{N}(0, 1)$. El objetivo de esta sección es proveer condiciones bajo las cuales variables aleatorias de la forma $U = \mathbf{X}^\top \mathbf{A} \mathbf{X}$ con $\mathbf{Y} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ siguen una distribución chi-cuadrado no central así como establecer condiciones para la independencia entre formas cuadráticas.

RESULTADO 1.68. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \mathbf{I})$ y $\mathbf{A} \in \mathbb{R}^{p \times p}$ es matriz simétrica. Entonces $\mathbf{X}^\top \mathbf{A} \mathbf{X} \sim \chi^2(k; \theta)$ si y sólo si \mathbf{A} es idempotente, en cuyo caso los grados de libertad y el parámetro de no centralidad están dados por

$$k = \text{rg}(\mathbf{A}) = \text{tr}(\mathbf{A}), \quad y \quad \theta = \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu},$$

respectivamente.

DEMOSTRACIÓN. Suponga que \mathbf{A} es idempotente de rango k . Entonces existe una matriz ortogonal \mathbf{P} tal que

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Sea $\mathbf{Y} = \mathbf{P}^\top \mathbf{X}$, entonces $\mathbf{Y} \sim \mathbf{N}_p(\mathbf{P}^\top \boldsymbol{\mu}, \mathbf{I})$, y

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{Y}^\top \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Y} = \sum_{i=1}^k Y_i^2,$$

que sigue una distribución chi-cuadrado con k grados de libertad. Para el parámetro de no centralidad θ , note que

$$\begin{aligned} \mathbb{E}\{\chi^2(k; \theta)\} &= k + 2\theta = \mathbb{E}(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \text{tr}(\mathbb{E}(\mathbf{X} \mathbf{X}^\top) \mathbf{A}) \\ &= \text{tr}((\mathbf{I} + \boldsymbol{\mu} \boldsymbol{\mu}^\top) \mathbf{A}) = k + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}, \end{aligned}$$

y de ahí que $\theta = \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$.

Ahora, suponga que $\mathbf{X}^\top \mathbf{A} \mathbf{X} \sim \chi^2(k; \theta)$. Si \mathbf{A} tiene rango r , entonces para \mathbf{P} matriz ortogonal $p \times p$,

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

con $\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_r)$, donde $\lambda_1, \dots, \lambda_r$ son los valores propios no nulos de \mathbf{A} . Sea $\mathbf{Y} = \mathbf{P}^\top \mathbf{X}$, entonces

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{Y}^\top \mathbf{P}^\top \mathbf{A} \mathbf{P} \mathbf{Y} = \sum_{j=1}^r \lambda_j Y_j^2 = U.$$

Tenemos que $\mathbf{Y} \sim \mathbf{N}_p(\boldsymbol{\delta}, \mathbf{I})$ con $\boldsymbol{\delta} = \mathbf{P}^\top \boldsymbol{\mu}$, de modo que $Y_j^2 \sim \chi^2(1; \delta_j^2/2)$ con función característica

$$\varphi_{Y_j^2}(t) = (1 - 2it)^{-1/2} \exp\left(\frac{it\delta_j^2}{1 - 2it}\right),$$

por la independencia de Y_1, \dots, Y_r sigue que

$$\begin{aligned} \varphi_U(t) &= \prod_{j=1}^r (1 - 2it\lambda_j)^{-1/2} \exp\left(\frac{it\lambda_j\delta_j^2}{1 - 2it\lambda_j}\right) \\ &= \exp\left(it \sum_{j=1}^r \frac{\lambda_j\delta_j^2}{1 - 2it\lambda_j}\right) \prod_{j=1}^r (1 - 2it\lambda_j)^{-1/2}. \end{aligned}$$

Como $\mathbf{X}^\top \mathbf{A} \mathbf{X} \sim \chi_k^2(\theta)$ tiene función característica

$$\varphi_{\mathbf{X}^\top \mathbf{A} \mathbf{X}}(t) = (1 - 2it)^{-k/2} \exp\left(\frac{2it\theta}{1 - 2it}\right),$$

entonces desde las dos expresiones anteriores debemos tener $r = k$, $\lambda_j = 1$, $\forall j$ y $\theta = \sum_j \delta_j^2/2$. Consecuentemente $\mathbf{P}^\top \mathbf{A} \mathbf{P}$ tiene la forma

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

que es idempotente. Luego

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = (\mathbf{P}^\top \mathbf{A} \mathbf{P})(\mathbf{P}^\top \mathbf{A} \mathbf{P}) = \mathbf{P}^\top \mathbf{A}^2 \mathbf{P} \implies \mathbf{A}^2 = \mathbf{A}.$$

□

RESULTADO 1.69. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ donde $\boldsymbol{\Sigma}$ es no singular y \mathbf{X} , $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son particionados 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 , $\boldsymbol{\mu}_1$ son $k \times 1$ y $\boldsymbol{\Sigma}_{11}$ es $k \times k$. Entonces

$$U = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) - (\mathbf{X}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu}_1) \sim \chi^2(p - k).$$

DEMOSTRACIÓN. Considere $\boldsymbol{\Sigma} = \mathbf{B} \mathbf{B}^\top$, donde \mathbf{B} es no singular y particione \mathbf{B} como

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}, \quad \mathbf{B}_1 \in \mathbb{R}^{k \times p}.$$

Luego,

$$\boldsymbol{\Sigma} = \mathbf{B} \mathbf{B}^\top = \begin{pmatrix} \mathbf{B}_1 \mathbf{B}_1^\top & \mathbf{B}_1 \mathbf{B}_2^\top \\ \mathbf{B}_2 \mathbf{B}_1^\top & \mathbf{B}_2 \mathbf{B}_2^\top \end{pmatrix},$$

de donde sigue que $\Sigma_{11} = B_1 B_1^\top$. Ahora, sea $Z = B^{-1}(X - \mu) \sim N_p(0, I)$. De este modo,

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} Z = \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}.$$

Entonces

$$\begin{aligned} U &= Z^\top Z - Z^\top B_1^\top (B_1 B_1^\top)^{-1} B_1 Z = Z^\top (I - B_1^\top (B_1 B_1^\top)^{-1} B_1) Z \\ &= Z^\top (I - H_1) Z, \quad \text{con } H_1 = B_1^\top (B_1 B_1^\top)^{-1} B_1. \end{aligned}$$

Note que H_1 es simétrica e idempotente y por tanto también lo es $C = I - H_1$. De donde sigue que $U \sim \chi^2(\nu)$, con $\nu = \text{rg}(C) = p - k$. \square

El Resultado 1.68 se puede generalizar al caso que X tiene una matriz de covarianza arbitraria. Suponga que $X \sim N_p(0, \Sigma)$. Una condición para que $X^\top A X$ tenga una distribución chi-cuadrado es

$$\Sigma A \Sigma A = \Sigma A,$$

en cuyo caso los grados de libertad son $k = \text{rg}(A \Sigma)$. Si Σ es no singular, la condición resulta $A \Sigma A = A$.

RESULTADO 1.70. Si $X \sim N_p(0, \Sigma)$ donde Σ tiene rango k ($\leq p$) y si A es una inversa generalizada de Σ ($\Sigma A \Sigma = \Sigma$), entonces $X^\top A X \sim \chi^2(k)$.

DEMOSTRACIÓN. Considere $Y = B X$ donde B es una matriz no singular $p \times p$ tal que

$$B \Sigma B^\top = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Particionando $Y = (Y_1^\top, Y_2^\top)^\top$ donde Y_1 es un vector $k \times 1$ sigue que $Y_1 \sim N_k(0, I)$ y $Y_2 = 0$ con probabilidad 1. Es decir, tenemos que

$$Y = (Y_1^\top, 0)^\top, \quad \text{con probabilidad 1.}$$

Ahora, note que

$$\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} = B \Sigma B^\top = B \Sigma A \Sigma B^\top$$

pues A es una inversa generalizada de Σ . De este modo,

$$\begin{aligned} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} &= B \Sigma B^\top B^{-\top} A B^{-1} B \Sigma B^\top \\ &= \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} B^{-\top} A B^{-1} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Luego, con probabilidad uno,

$$\begin{aligned} X^\top A X &= Y^\top B^{-\top} A B^{-1} Y = (Y_1^\top, 0) B^{-\top} A B^{-1} \begin{pmatrix} Y_1 \\ 0 \end{pmatrix} \\ &= (Y_1^\top, 0) \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} B^{-\top} A B^{-1} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ 0 \end{pmatrix} \\ &= (Y_1^\top, 0) \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ 0 \end{pmatrix} = Y_1^\top Y_1 \sim \chi^2(k). \end{aligned}$$

\square

RESULTADO 1.71. Si $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\Sigma}$ es no singular, y \mathbf{A} es una matriz simétrica $p \times p$. Entonces $\mathbf{X}^\top \mathbf{A} \mathbf{X} \sim \chi^2(k; \lambda)$, donde $k = \text{rg}(\mathbf{A})$, $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} / 2$ si y sólo si $\mathbf{A} \boldsymbol{\Sigma}$ es matriz idempotente.

DEMOSTRACIÓN. Considere $\mathbf{Y} = \mathbf{B} \mathbf{X}$, donde \mathbf{B} es una matriz no singular $p \times p$ tal que $\mathbf{B} \boldsymbol{\Sigma} \mathbf{B}^\top = \mathbf{I}_p$. Entonces

$$\mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{Y}^\top \mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1} \mathbf{Y},$$

donde $\mathbf{Y} \sim \mathbf{N}_p(\mathbf{B} \boldsymbol{\mu}, \mathbf{I})$. Desde el Resultado 1.68 sigue que $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ tiene distribución chi-cuadrado sólo si $\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1}$ es idempotente. Esto es equivalente a mostrar que $\mathbf{A} \boldsymbol{\Sigma}$ es idempotente.

Si $\mathbf{A} \boldsymbol{\Sigma}$ es idempotente, tenemos

$$\mathbf{A} = \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} = \mathbf{A} \mathbf{B}^{-1} \mathbf{B}^{-\top} \mathbf{A}, \quad (\boldsymbol{\Sigma} = \mathbf{B}^{-1} \mathbf{B}^{-\top})$$

así, pre- y post-multiplicando por $\mathbf{B}^{-\top}$ y \mathbf{B}^{-1} , obtenemos

$$\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1} = (\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1})(\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1}),$$

y por tanto es idempotente.

Por otro lado, si $\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1}$ es idempotente, entonces

$$\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1} = (\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1})(\mathbf{B}^{-\top} \mathbf{A} \mathbf{B}^{-1}) = \mathbf{B}^{-\top} \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \mathbf{B}^{-1},$$

es decir $\mathbf{A} = \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}$ y de ahí que $\mathbf{A} \boldsymbol{\Sigma}$ es idempotente. \square

EJEMPLO 1.72. Sea X_1, \dots, X_n variables aleatorias IID $\mathbf{N}(\theta, \sigma^2)$, en este caso podemos definir $\mathbf{X} = (X_1, \dots, X_n)^\top$ tal que $\mathbf{X} \sim \mathbf{N}_n(\theta \mathbf{1}_n, \sigma^2 \mathbf{I})$. Considere la forma cuadrática

$$Q = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{C} \mathbf{X} = \mathbf{X}^\top \mathbf{A} \mathbf{X},$$

con $\mathbf{C} = \mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^\top$ y $\mathbf{A} = \mathbf{C} / \sigma^2$. De esta manera

$$\mathbf{A} \boldsymbol{\Sigma} = \mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^\top,$$

que es idempotente. Además

$$\text{rg}(\mathbf{A}) = \text{tr}(\mathbf{C}) = \text{tr}\left(\mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^\top\right) = n - 1,$$

y

$$\lambda = \frac{\theta^2}{2} \mathbf{1}^\top \mathbf{A} \mathbf{1} = \frac{\theta^2}{2\sigma^2} \mathbf{1}^\top \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^\top\right) \mathbf{1} = 0.$$

Finalmente,

$$Q = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n - 1).$$

RESULTADO 1.73. Sea $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $Q_1 = \mathbf{X}^\top \mathbf{A} \mathbf{X}$ y $Q_2 = \mathbf{X}^\top \mathbf{B} \mathbf{X}$. Entonces Q_1 y Q_2 son independientes si y sólo si $\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} = \mathbf{0}$.

DEMOSTRACIÓN. Tenemos $\boldsymbol{\Sigma} = \mathbf{T} \mathbf{T}^\top$, y defina $\mathbf{G}_1 = \mathbf{T}^\top \mathbf{A} \mathbf{T}$, $\mathbf{G}_2 = \mathbf{T}^\top \mathbf{B} \mathbf{T}$. Note que si $\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} = \mathbf{0}$, entonces

$$\mathbf{G}_1 \mathbf{G}_2 = (\mathbf{T}^\top \mathbf{A} \mathbf{T})(\mathbf{T}^\top \mathbf{B} \mathbf{T}) = \mathbf{T}^\top \mathbf{A} \boldsymbol{\Sigma} \mathbf{B} \mathbf{T} = \mathbf{0}.$$

Debido a la simetría de \mathbf{G}_1 y \mathbf{G}_2 , sigue que

$$\mathbf{0} = (\mathbf{G}_1 \mathbf{G}_2)^\top = \mathbf{G}_2^\top \mathbf{G}_1^\top = \mathbf{G}_2 \mathbf{G}_1.$$

Como $\mathbf{G}_1 \mathbf{G}_2 = \mathbf{G}_2 \mathbf{G}_1$ existe una matriz ortogonal \mathbf{P} que simultáneamente diagonaliza \mathbf{G}_1 y \mathbf{G}_2 , esto es:

$$\mathbf{P}^\top \mathbf{G}_1 \mathbf{P} = \mathbf{P}^\top \mathbf{T}^\top \mathbf{A} \mathbf{P} = \mathbf{D}_1,$$

$$\mathbf{P}^\top \mathbf{G}_2 \mathbf{P} = \mathbf{P}^\top \mathbf{T}^\top \mathbf{B} \mathbf{P} = \mathbf{D}_2.$$

De este modo,

$$\mathbf{0} = \mathbf{G}_1 \mathbf{G}_2 = \mathbf{P} \mathbf{D}_1 \mathbf{P}^\top \mathbf{P} \mathbf{D}_2 \mathbf{P}^\top = \mathbf{P} \mathbf{D}_1 \mathbf{D}_2 \mathbf{P}^\top$$

lo que es verdad si $\mathbf{D}_1 \mathbf{D}_2 = \mathbf{0}$. Como \mathbf{D}_1 y \mathbf{D}_2 son diagonales, sus elementos diagonales deben ocurrir en posiciones diferentes. Es decir, podemos escribir

$$\mathbf{D}_1 = \begin{pmatrix} M_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M_2 \end{pmatrix}.$$

Sea $\mathbf{Y} = \mathbf{P}^\top \mathbf{T}^{-1} \mathbf{X}$, entonces

$$Q_1 = \mathbf{X}^\top \mathbf{A} \mathbf{X} = \mathbf{X}^\top \mathbf{T}^{-\top} \mathbf{P} \mathbf{P}^\top \mathbf{T}^\top \mathbf{A} \mathbf{P} \mathbf{P}^\top \mathbf{T}^{-1} \mathbf{X} = \mathbf{Y}^\top \mathbf{D}_1 \mathbf{Y},$$

$$Q_2 = \mathbf{X}^\top \mathbf{B} \mathbf{X} = \mathbf{X}^\top \mathbf{T}^{-\top} \mathbf{P} \mathbf{P}^\top \mathbf{T}^\top \mathbf{B} \mathbf{P} \mathbf{P}^\top \mathbf{T}^{-1} \mathbf{X} = \mathbf{Y}^\top \mathbf{D}_2 \mathbf{Y}.$$

Además,

$$\text{Cov}(\mathbf{Y}) = \text{Cov}(\mathbf{P}^\top \mathbf{T}^{-1} \mathbf{X}) = \mathbf{P}^\top \mathbf{T}^{-1} \text{Cov}(\mathbf{X}) \mathbf{T}^{-\top} \mathbf{P} = \mathbf{I}.$$

En efecto, $\mathbf{Y} \sim \mathbf{N}_p(\mathbf{P}^\top \mathbf{T}^{-1} \boldsymbol{\mu}, \mathbf{I})$. Ahora, particionando adecuadamente \mathbf{Y} , sigue que

$$\mathbf{Y}^\top \mathbf{D}_1 \mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top) \begin{pmatrix} M_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \mathbf{Y}_1^\top M_1 \mathbf{Y}_1,$$

$$\mathbf{Y}^\top \mathbf{D}_2 \mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M_2 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \mathbf{Y}_2^\top M_2 \mathbf{Y}_2,$$

y la independencia entre Q_1 y Q_2 sigue desde la independencia entre \mathbf{Y}_1 y \mathbf{Y}_2 . \square

RESULTADO 1.74. Sea $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $Q = \mathbf{X}^\top \mathbf{A} \mathbf{X}$ y $U = \mathbf{B} \mathbf{X}$. Entonces Q y U son independientes si y sólo si $\mathbf{B} \boldsymbol{\Sigma} \mathbf{A} = \mathbf{0}$.

EJEMPLO 1.75. Considere X_1, \dots, X_n muestra aleatoria desde $\mathbf{N}(\theta, \sigma^2)$, así

$$\mathbf{X} = (X_1, \dots, X_n)^\top \sim \mathbf{N}_n(\theta \mathbf{1}, \sigma^2 \mathbf{I}_n).$$

Tenemos

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \mathbf{1}^\top \mathbf{X}, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \mathbf{X}^\top \mathbf{C} \mathbf{X}.$$

Como $\mathbf{C} \mathbf{1} = \mathbf{0}$ sigue la independencia entre \bar{X} y S^2 .

Considere los siguientes dos lemas, los que permitirán mostrar el resultado principal de esta sección.

LEMA 1.76. Sean $\mathbf{A}_1, \dots, \mathbf{A}_k$ matrices $m \times m$ simétrica e idempotente y suponga que

$$\mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_k = \mathbf{I}_m.$$

Entonces $\mathbf{A}_i \mathbf{A}_j = \mathbf{0}$ para todo $i \neq j$.

DEMOSTRACIÓN. Considere cualquiera de esas matrices, digamos \mathbf{A}_h y denote su rango por r . Como \mathbf{A}_h es simétrica e idempotente, existe una matriz ortogonal \mathbf{P} tal que

$$\mathbf{P}^\top \mathbf{A}_h \mathbf{P} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Para $j \neq h$, defina $\mathbf{B}_j = \mathbf{P}^\top \mathbf{A}_j \mathbf{P}$, y note que

$$\mathbf{I}_m = \mathbf{P}^\top \mathbf{P} = \mathbf{P}^\top \left(\sum_{j=1}^k \mathbf{A}_j \right) \mathbf{P} = \sum_{j=1}^k \mathbf{P}^\top \mathbf{A}_j \mathbf{P} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \sum_{j \neq h} \mathbf{B}_j.$$

O equivalentemente,

$$\sum_{j \neq h} \mathbf{B}_j = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-r} \end{pmatrix}.$$

Claramente, dado que \mathbf{A}_j es simétrica e idempotente, sigue que \mathbf{B}_j también lo es. De modo que, sus elementos diagonales son no negativos. Además, $(\mathbf{B}_j)_{ll} = 0$, para $l = 1, \dots, r$. Así, sigue que \mathbf{B}_j debe ser de la forma

$$\mathbf{B}_j = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_j \end{pmatrix},$$

donde \mathbf{C}_j es matriz $(m-r) \times (m-r)$, simétrica e idempotente. Ahora, para cualquier $j \neq h$

$$\mathbf{P}^\top \mathbf{A}_h \mathbf{A}_j \mathbf{P} = (\mathbf{P}^\top \mathbf{A}_h \mathbf{P})(\mathbf{P}^\top \mathbf{A}_j \mathbf{P}) = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_j \end{pmatrix} = \mathbf{0},$$

lo que es verdad, sólo si $\mathbf{A}_h \mathbf{A}_j = \mathbf{0}$, pues \mathbf{P} es no singular. Notando que h es arbitrario, la prueba es completa. \square

LEMA 1.77. Sean $\mathbf{A}_1, \dots, \mathbf{A}_k$ matrices simétricas de orden $m \times m$ y defina

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_k.$$

Considere las siguientes afirmaciones,

- (a) \mathbf{A}_i es idempotente, para $i = 1, \dots, k$.
- (b) \mathbf{A} es idempotente.
- (c) $\mathbf{A}_i \mathbf{A}_j = \mathbf{0}$, para $i \neq j$.

Entonces, si dos condiciones son satisfechas, la tercera condición debe ser verdadera.

DEMOSTRACIÓN. Primero mostraremos que (a) y (b) implica (c). Como \mathbf{A} es simétrica e idempotente, existe una matriz ortogonal \mathbf{P} tal que

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = \mathbf{P}^\top (\mathbf{A}_1 + \dots + \mathbf{A}_k) \mathbf{P} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (1.7)$$

donde $r = \text{rg}(\mathbf{A})$.

Sea $\mathbf{B}_i = \mathbf{P}^\top \mathbf{A}_i \mathbf{P}$, para $i = 1, \dots, k$, y note que \mathbf{B}_i es simétrica e idempotente. Es decir, \mathbf{B}_i debe ser de la forma

$$\mathbf{B}_i = \begin{pmatrix} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

donde la matriz $r \times r$, \mathbf{C}_i debe ser simétrica e idempotente. Por (1.7), tenemos

$$\mathbf{C}_1 + \dots + \mathbf{C}_k = \mathbf{I}_r.$$

Por el Lema 1.76, sigue que $\mathbf{C}_i \mathbf{C}_j = \mathbf{0}$ para $i \neq j$, de donde obtenemos $\mathbf{B}_i \mathbf{B}_j = \mathbf{0}$ y de ahí que $\mathbf{A}_i \mathbf{A}_j = \mathbf{0}$, para $i \neq j$.

(a) y (c) implican (b), sigue de notar

$$\begin{aligned} \mathbf{A}^2 &= \left(\sum_{i=1}^k \mathbf{A}_i \right)^2 = \sum_{i=1}^k \sum_{j=1}^k \mathbf{A}_i \mathbf{A}_j = \sum_{i=1}^k \mathbf{A}_i^2 + \sum_{i \neq j} \sum \mathbf{A}_i \mathbf{A}_j \\ &= \sum_{i=1}^k \mathbf{A}_i = \mathbf{A}. \end{aligned}$$

Finalmente, para probar que (b) y (c) implican (a). Suponga que (c) es verdad, entonces $\mathbf{A}_i \mathbf{A}_j = \mathbf{A}_j \mathbf{A}_i$ para todo $i \neq j$ y las matrices $\mathbf{A}_1, \dots, \mathbf{A}_k$ pueden ser diagonalizadas simultáneamente. Esto es, existe una matriz ortogonal \mathbf{Q} tal que

$$\mathbf{Q}^\top \mathbf{A}_i \mathbf{Q} = \mathbf{D}_i, \quad i = 1, \dots, k,$$

donde cada una de las matrices $\mathbf{D}_1, \dots, \mathbf{D}_k$ es diagonal. Además,

$$\mathbf{D}_i \mathbf{D}_j = \mathbf{Q}^\top \mathbf{A}_i \mathbf{Q} \mathbf{Q}^\top \mathbf{A}_j \mathbf{Q} = \mathbf{Q}^\top \mathbf{A}_i \mathbf{A}_j \mathbf{Q} = \mathbf{0}, \quad i \neq j. \quad (1.8)$$

Como \mathbf{A} es simétrica e idempotente, también lo es la matriz diagonal

$$\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \mathbf{D}_1 + \dots + \mathbf{D}_k,$$

y cada elemento diagonal de $\mathbf{Q}^\top \mathbf{A} \mathbf{Q}$ debe ser 0 o 1, y por (1.8), lo mismo es válido para los elementos diagonales de $\mathbf{D}_1, \dots, \mathbf{D}_k$.

De este modo, \mathbf{D}_i es simétrica e idempotente y de ahí que también lo es

$$\mathbf{A}_i = \mathbf{Q} \mathbf{D}_i \mathbf{Q}^\top, \quad i = 1, \dots, k,$$

lo que termina la prueba. \square

OBSERVACIÓN 1.78. Suponga que las condiciones del Lema son satisfechas. Entonces (a) implica que $\text{rg}(\mathbf{A}_i) = \text{tr}(\mathbf{A}_i)$, y desde (b), sigue que

$$\text{rg}(\mathbf{A}) = \text{tr}(\mathbf{A}) = \text{tr} \left(\sum_{i=1}^k \mathbf{A}_i \right) = \sum_{i=1}^k \text{tr}(\mathbf{A}_i) = \sum_{i=1}^k \text{rg}(\mathbf{A}_i).$$

RESULTADO 1.79 (Teorema de Cochran). Sea $\mathbf{X} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con $\boldsymbol{\Sigma} > \mathbf{0}$. Suponga que \mathbf{A}_i , es una matriz simétrica de orden $p \times p$ con rango r_i , para $i = 1, \dots, k$, y

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_k,$$

es de rango r . Considere las condiciones:

- (a) $\mathbf{A}_i \boldsymbol{\Sigma}$ es idempotente, para $i = 1, \dots, k$.
- (b) $\mathbf{A} \boldsymbol{\Sigma}$ es idempotente.
- (c) $\mathbf{A}_i \boldsymbol{\Sigma} \mathbf{A}_j = \mathbf{0}$, para $i \neq j$.
- (d) $r = \sum_{i=1}^k r_i$.

si dos de (a), (b) y (c) se satisfacen, o si (b) y (d) son satisfechas. Entonces,

- (i) $\mathbf{X}^\top \mathbf{A}_i \mathbf{X} \sim \chi^2(r_i; \lambda_i)$, con $\lambda_i = \boldsymbol{\mu}^\top \mathbf{A}_i \boldsymbol{\mu} / 2$, para $i = 1, \dots, k$.
- (ii) $\mathbf{X}^\top \mathbf{A} \mathbf{X} \sim \chi^2(r; \lambda)$, con $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} / 2$.
- (iii) $\mathbf{X}^\top \mathbf{A}_1 \mathbf{X}, \mathbf{X}^\top \mathbf{A}_2 \mathbf{X}, \dots, \mathbf{X}^\top \mathbf{A}_k \mathbf{X}$ son mutuamente independientes.

DEMOSTRACIÓN. Tenemos que $\Sigma = \mathbf{T}\mathbf{T}^\top$ y las condiciones (a)-(d), pueden ser expresadas como:

- (a) $\mathbf{T}^\top \mathbf{A}_i \mathbf{T}$ es idempotente, para $i = 1, \dots, k$.
- (b) $\mathbf{T}^\top \mathbf{A} \mathbf{T}$ es idempotente.
- (c) $(\mathbf{T}^\top \mathbf{A}_i \mathbf{T})(\mathbf{T}^\top \Sigma \mathbf{A}_j \mathbf{T}) = \mathbf{0}$, para $i \neq j$.
- (d) $\text{rg}(\mathbf{T}^\top \mathbf{A} \mathbf{T}) = \sum_{i=1}^k \text{rg}(\mathbf{T}^\top \mathbf{A}_i \mathbf{T})$.

Como $\mathbf{T}^\top \mathbf{A}_1 \mathbf{T}, \mathbf{T}^\top \mathbf{A}_2 \mathbf{T}, \dots, \mathbf{T}^\top \mathbf{A}_k \mathbf{T}$ y $\mathbf{T}^\top \mathbf{A} \mathbf{T}$ satisfacen las condiciones del Lema 1.78 (y de la Observación 1.78). Entonces, las condiciones (a)-(d) se satisfacen.

Sabemos que (a) implica (i) y (b) implica (ii). Mientras que, Resultado 1.73 con (c), garantiza (iii), lo que completa la prueba. \square

Ejercicios

- 1.1 Sean $\mathbf{X}_1, \dots, \mathbf{X}_n$ vectores aleatorios independientes con $\mathbf{X}_i \sim \mathbf{N}_p(\boldsymbol{\mu}, \Sigma)$, para $i = 1, \dots, n$. Obtenga la distribución de

$$\sum_{i=1}^n \alpha_i \mathbf{X}_i,$$

con $\alpha_1, \dots, \alpha_n$ constantes fijas.

- 1.2 Si $\mathbf{X}_1, \dots, \mathbf{X}_n$ son independientes cada uno con $\mathbf{X}_i \sim \mathbf{N}_p(\boldsymbol{\mu}, \Sigma)$. Muestre que la distribución del vector de medias

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

es $\mathbf{N}_p(\boldsymbol{\mu}, \frac{1}{n} \Sigma)$.

- 1.3 Demuestre el Resultado 1.20, usando la función característica de un vector aleatorio normal.

- 1.4 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas $\mathbf{N}(\mu, \sigma^2)$ y defina

$$Q = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2,$$

¿Es Q un estimador insesgado de σ^2 ?

- 1.5 Sea $\mathbf{X} \sim \mathbf{N}_n(\boldsymbol{\mu}, \Sigma)$ y defina

$$\mathbf{Y} = \mathbf{T}^\top \Sigma^{-1/2} (\mathbf{X} - \boldsymbol{\mu}), \quad \mathbf{u} = \mathbf{T}^\top \Sigma^{1/2} \mathbf{A} \boldsymbol{\mu}.$$

con \mathbf{T} ortogonal y $\mathbf{A} = \mathbf{A}^\top$. Obtenga la distribución de \mathbf{Y} y calcule $\text{var}(\mathbf{u}^\top \mathbf{Y})$.

- 1.6 Considere \mathbf{Z} matriz aleatoria $n \times p$ con función característica

$$\varphi_{\mathbf{Z}}(\mathbf{T}) = \mathbf{E}\{\exp(i \text{tr}(\mathbf{T}^\top \mathbf{Z}))\} = \exp\{-\frac{1}{2} \text{tr}(\mathbf{T}^\top \mathbf{T})\}.$$

con $\mathbf{T} \in \mathbb{R}^{n \times p}$. Obtenga la función característica de

$$\mathbf{Y} = \Sigma^{1/2} \mathbf{Z} \Theta^{1/2} + \boldsymbol{\mu},$$

donde $\boldsymbol{\mu} \in \mathbb{R}^{n \times p}$ y $\boldsymbol{\Sigma}$, $\boldsymbol{\Theta}$ son matrices semidefinidas positivas $n \times n$ y $p \times p$, respectivamente.

- 1.7 Sea $\mathbf{Z} = \mathbf{U}\mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ con $\mathbf{U} \in \mathbb{R}^{n \times p}$ tal que $\mathbf{U}^\top \mathbf{U} = \mathbf{I}$, \mathbf{D} es matriz diagonal $p \times p$ y $\boldsymbol{\epsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$. Considere

$$\hat{\boldsymbol{\alpha}} = (\mathbf{D}^2 + \lambda \mathbf{I})^{-1} \mathbf{D} \mathbf{U}^\top \mathbf{Z}.$$

donde λ es un escalar positivo.

- (a) Obtenga la distribución de $\hat{\boldsymbol{\alpha}}$,
 (b) Muestre que

$$\boldsymbol{\alpha} - \mathbf{E}(\hat{\boldsymbol{\alpha}}) = \lambda(\mathbf{D}^2 + \lambda \mathbf{I})^{-1} \boldsymbol{\alpha}.$$

- 1.8 Sea $\mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ y considere $\mathbf{b} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$, $\mathbf{u} = (\mathbf{D}^{-1} + \mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top (\mathbf{Y} - \mathbf{X}\mathbf{b})$, donde $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{Z} \in \mathbb{R}^{n \times q}$ y \mathbf{D} es matriz no singular $q \times q$.

- (a) Halle la distribución de \mathbf{b} y \mathbf{u} ,
 (b) ¿Son \mathbf{b} y \mathbf{u} independientes?

- 1.9 Considere

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{b} \end{pmatrix} \sim \mathbf{N}_{n+q} \left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}\mathbf{D}\mathbf{Z}^\top + \mathbf{R} & \mathbf{Z}\mathbf{D} \\ \mathbf{D}\mathbf{Z}^\top & \mathbf{D} \end{pmatrix} \right),$$

donde $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{Z} \in \mathbb{R}^{n \times q}$ y \mathbf{R} , \mathbf{D} son matrices no singulares $n \times n$ y $q \times q$, respectivamente. Determine la distribución de $\mathbf{b}|\mathbf{Y}$.

- 1.10 Sea $U_i \sim \chi^2(n_i; \lambda_i)$, $i = 1, \dots, K$ variables aleatorias independientes. Muestre que

$$U = \sum_{i=1}^K U_i \sim \chi^2(n; \lambda),$$

donde $n = \sum_{i=1}^K n_i$ y $\lambda = \sum_{i=1}^K \lambda_i$.

- 1.11 Sea $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$, donde \mathbf{X} es matriz $n \times p$ con $\text{rg}(\mathbf{X}) = p$ y $\mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$. Defina

$$Q = \frac{(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})^\top [\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top]^{-1} (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})}{\sigma^2},$$

donde $\mathbf{G} \in \mathbb{R}^{m \times p}$ con $\text{rg}(\mathbf{G}) = m$ y \mathbf{g} es vector m -dimensional. Determine la distribución de Q .

- 1.12 Sea $\mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ y considere las formas cuadráticas

$$Q_1 = \frac{\hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}}}{\sigma^2}, \quad Q_2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2},$$

donde $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ con $\mathbf{X} \in \mathbb{R}^{n \times p}$ y $\text{rg}(\mathbf{X}) = p$.

- (a) Halle la distribución de Q_i , $i = 1, 2$.
 (b) Sea $Q = Q_1 + Q_2$, mostrar la independencia conjunta de Q_1 y Q_2 .

- 1.13 Sea $\mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ con $\mathbf{X} \in \mathbb{R}^{n \times p}$ y $\boldsymbol{\beta} \in \mathbb{R}^p$. Muestre que

$$Q = \frac{\mathbf{Y}^\top (\mathbf{I} - \frac{1}{n} \mathbf{J}) \mathbf{Y}}{\sigma^2}, \quad Q_1 = \frac{\mathbf{Y}^\top (\mathbf{I} - \mathbf{H}) \mathbf{Y}}{\sigma^2}, \quad Q_2 = \frac{\mathbf{Y}^\top (\mathbf{H} - \frac{1}{n} \mathbf{J}) \mathbf{Y}}{\sigma^2},$$

con $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$, tienen distribuciones chi-cuadrado independientes.

1.14 Considere

$$\mathbf{Y} = (\mathbf{I}_p \otimes \mathbf{1}_n) \boldsymbol{\alpha} + \boldsymbol{\epsilon},$$

donde $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top, \dots, \mathbf{Y}_p^\top)^\top$ con \mathbf{Y}_i vector n -dimensional, para $i = 1, \dots, p$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$ y $\boldsymbol{\epsilon} \sim \mathbf{N}_{np}(\mathbf{0}, \sigma^2 \mathbf{I}_{np})$. Sean

$$Q_1 = \frac{\mathbf{Y}^\top (\mathbf{I}_p \otimes \frac{1}{n} \mathbf{J}_n) \mathbf{Y}}{\sigma^2}, \quad \text{y} \quad Q_2 = \frac{\mathbf{Y}^\top (\mathbf{I}_p \otimes \mathbf{C}) \mathbf{Y}}{\sigma^2},$$

donde $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\top$ y $\mathbf{C} = \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n$.

(a) Halle la distribución de Q_k , $k = 1, 2$.

(b) ¿Son Q_1 y Q_2 independientes?