Capítulo 1

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 $\boldsymbol{X}$ es una función (medible) desde el espacio de probabilidad Ω a $\mathbb{R}^n$, esto es

$$\boldsymbol{X} : \Omega \to \mathbb{R}^n.$$

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

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

$$F(\boldsymbol{x}) = \mathsf{P}(\boldsymbol{X} \leq \boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^n \tag{1.1}$$

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

La función en (1.1) debe ser entendida como

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Naturalmente la función de densidad debe satisfacer

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

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

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

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

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

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

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

$$f_{X_2|X_1=x_1}(\boldsymbol{u}) = \frac{f_X(\boldsymbol{x}_1, \boldsymbol{u})}{f_1(\boldsymbol{x}_1)},$$

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

$$F_{X_2|X_1=x_1}(\boldsymbol{u}) = \int_{-\infty}^{\boldsymbol{u}} f_{X_2|X_1=x_1}(\boldsymbol{t})\, \mathsf{d}\boldsymbol{t},$$

tenemos además que

$$f_{X_2|X_1=x_1}(\boldsymbol{u}) = \frac{f_X(\boldsymbol{x}_1, \boldsymbol{u})}{\int_{\mathbb{R}^{n_2}} f_X(\boldsymbol{x}_1, \boldsymbol{t})\, \mathsf{d}\boldsymbol{t}}.$$

1.2. Operadores de esperanza y covarianza

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

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

siempre que la integral (n-dimensional) exista.

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

$$\mathsf{E}(\boldsymbol{Z}(\boldsymbol{X})) = (\mathsf{E}(Z_{ij})), \qquad Z_{ij} = Z_{ij}(\boldsymbol{X}). \tag{1.2}$$

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

$$\mathsf{E}(\boldsymbol{A}) = \boldsymbol{A}.$$

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

$$\mathsf{E}(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A}\,\mathsf{E}(\boldsymbol{Z})\boldsymbol{B} + \boldsymbol{C}.$$

Demostración. Sea $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z}\boldsymbol{B} + \boldsymbol{C}$, entonces

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

de este modo

$$\begin{aligned}\mathsf{E}(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{B} + \boldsymbol{C}) = (\mathsf{E}(Y_{ij})) &= \Big(\sum_{r=1}^{m}\sum_{s=1}^{n} a_{ir}\,\mathsf{E}(Z_{rs}) b_{sj} + c_{ij}\Big)\\ &= \boldsymbol{A}\,\mathsf{E}(\boldsymbol{Z})\boldsymbol{B} + \boldsymbol{C}.\end{aligned}$$

□

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

$$\mathsf{E}\Big(\sum_i \boldsymbol{A}_i \boldsymbol{X}_i\Big) = \sum_i \boldsymbol{A}_i\,\mathsf{E}(\boldsymbol{X}_i),$$

de manera similar tenemos que

$$\mathsf{E}\Big(\sum_i \alpha_i \boldsymbol{Z}_i\Big) = \sum_i \alpha_i\,\mathsf{E}(\boldsymbol{Z}_i),$$

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

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

$$\mathsf{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = (\mathsf{Cov}(X_i, Y_j)).$$

Podemos apreciar, a partir de la definición de covarianza que

$$\mathsf{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathsf{E}\{(\boldsymbol{X} - \mathsf{E}(\boldsymbol{X}))(\boldsymbol{Y} - \mathsf{E}(\boldsymbol{Y}))^\top\}.$$

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

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

1.3. Independencia de vectores aleatorios

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

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

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

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

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

o bien,

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

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

$$\mathsf{E}\{\kappa(\boldsymbol{X})\tau(\boldsymbol{Y})\} = \mathsf{E}\{\kappa(\boldsymbol{X})\}\,\mathsf{E}\{\tau(\boldsymbol{Y})\},$$

si las esperanzas existen.

Demostración. En efecto, es fácil notar que

$$\begin{aligned}
\mathsf{E}\{\kappa(\boldsymbol{X})\tau(\boldsymbol{Y})\} &= \int\int \kappa(\boldsymbol{x})\tau(\boldsymbol{y})g(\boldsymbol{x})h(\boldsymbol{y})\,\mathsf{d}\boldsymbol{x}\,\mathsf{d}\boldsymbol{y} \\
&= \Big(\int \kappa(\boldsymbol{x})g(\boldsymbol{x})\,\mathsf{d}\boldsymbol{x}\Big)\Big(\int \tau(\boldsymbol{y})h(\boldsymbol{y})\,\mathsf{d}\boldsymbol{y}\Big) \\
&= \mathsf{E}\{\kappa(\boldsymbol{X})\}\,\mathsf{E}\{\tau(\boldsymbol{Y})\}.
\end{aligned}$$

□

1.4. Cambios de variable

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

$$J(\boldsymbol{y} \to \boldsymbol{x}) = |\mathsf{D}\boldsymbol{f}(\boldsymbol{x})|_+ = \operatorname{abs}(\det(\mathsf{D}\boldsymbol{f}(\boldsymbol{x}))),$$

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

$$\begin{aligned}
J(\boldsymbol{z} \to \boldsymbol{x}) &= J(\boldsymbol{z} \to \boldsymbol{y}) \cdot J(\boldsymbol{y} \to \boldsymbol{x}) \\
J(\boldsymbol{y} \to \boldsymbol{x}) &= \{J(\boldsymbol{x} \to \boldsymbol{y})\}^{-1}
\end{aligned}$$

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 $\boldsymbol{X}$ vector aleatorio n-dimensional con densidad $f_X(\boldsymbol{x})$ y soporte $S = \{\boldsymbol{x} : f_X(\boldsymbol{x}) > 0\}$. Para $\boldsymbol{g} : S \to \mathbb{R}^n$ diferenciable e invertible, sea $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$. Entonces la densidad de $\boldsymbol{Y}$ está dada por*

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

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

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

vectorizando obtenemos

$$\operatorname{vec}\mathsf{d}\,\boldsymbol{Y} = (\boldsymbol{B}^\top \otimes \boldsymbol{A})\operatorname{vec}\mathsf{d}\,\boldsymbol{X},$$

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

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

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

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

en cuyo caso escribimos $Z \sim \mathsf{N}(0,1)$. 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 $Y \stackrel{\mathsf{d}}{=} \mu + \sigma Z$, donde $Z \sim \mathsf{N}(0,1)$. Cuando $\sigma^2 = 0$, la distribución $\mathsf{N}(\mu,\sigma^2)$ se interpreta como una distribución degenerada en μ. Si $Y \sim \mathsf{N}(\mu,\sigma^2)$, entonces su función característica adopta la forma

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

Sea $Z_1,\dots,Z_n$ variables aleatorias independientes cada una con distribución $\mathsf{N}(0,1)$ y considere el vector aleatorio $\boldsymbol{Z} = (Z_1,\dots,Z_n)^\top$. Entonces, es directo que $\mathsf{E}(\boldsymbol{Z}) = \boldsymbol{0}$ y $\mathsf{Cov}(\boldsymbol{Z}) = \boldsymbol{I}$.

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

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

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

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

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

por la Definición 1.18 tenemos que $\boldsymbol{h}^\top\boldsymbol{X}$ es normal y como c es una constante, sigue que $\boldsymbol{t}^\top\boldsymbol{Y}$ tiene distribución normal multivariada. □

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

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

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

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

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

$$\varphi_X(\boldsymbol{t}) = \mathsf{E}\{\exp(i\boldsymbol{t}^\top \boldsymbol{X})\} = \varphi_{t^\top X}(1),$$

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

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

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

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

Considere

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

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

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

□

OBSERVACIÓN 1.22. El Resultado 1.20 puede ser demostrado de manera bastante simple usando la función característica (ver Ejercicio 1.4).

RESULTADO 1.23. *Si* $\boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$. *Entonces*

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

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

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

y

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

de ahí que

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

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

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

□

Observación 1.24. Desde los Resultados 1.20 y 1.23, tenemos que para la variable aleatoria

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}, \qquad \boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I}),$$

sigue que $\mathsf{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ y $\mathsf{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$.

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

Demostración. Considere la siguiente partición:

$$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}, \tag{1.3}$$

donde $\boldsymbol{X}_1$ y $\boldsymbol{\mu}_1$ son vectores $k \times 1$ y $\boldsymbol{\Sigma}_{11}$ es $k \times k$. Aplicando el Resultado 1.20 con

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

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

Una consecuencia de este resultado es que la distribución marginal de *cada* componente de $\boldsymbol{X}$ 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 $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\boldsymbol{X}$, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son particionadas como en la Ecuación (1.3). Entonces los vectores $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ son independientes sólo si $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$.*

Demostración. Note que $\mathsf{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\Sigma}_{12}$, así la independencia entre $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ implica que $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$. Suponga ahora que $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$. Entonces la función característica

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

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

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

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

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

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

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

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

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

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

obtenemos

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

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

Ejemplo 1.29. Sea $\boldsymbol{X} \sim \mathsf{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma})$ donde

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad 0 \le \rho \le 1.$$

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

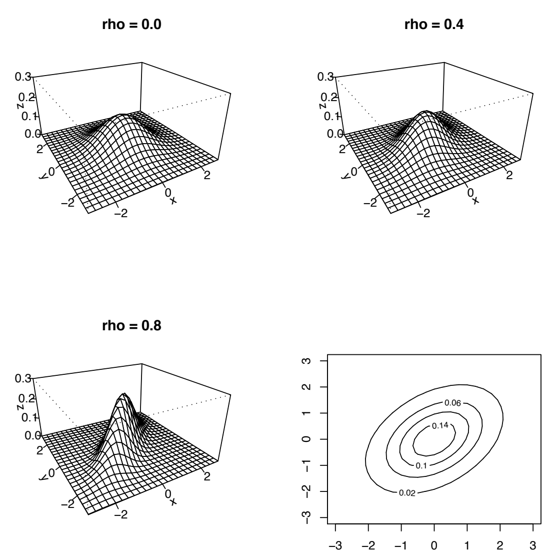

Figura 1. Función de densidad de $\boldsymbol{X} \sim \mathsf{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma})$ para $\rho = 0.0, 0.4$ y 0.8.

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

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

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

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

sigue una distribución chi-cuadrado con p grados de libertad y la cantidad $D = \{(\boldsymbol{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^{1/2}$ se conoce como *distancia de Mahalanobis* de $\boldsymbol{X}$ a $\boldsymbol{\mu}$.

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

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

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

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

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

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

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

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

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

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

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

Resultado 1.31. *Sea $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y particione $\boldsymbol{X}$, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ como:*

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

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

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

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

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

(b) *La distribución condicional*

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

Demostración. Considere la transformación lineal

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

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

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

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

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

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

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

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

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

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

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

Como

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

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

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

donde

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

lo que muestra el resultado. □

Observación 1.32. La esperanza de la distribución condicional de $\boldsymbol{X}_1$ dado $\boldsymbol{X}_2$, es decir

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

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

RESULTADO 1.33. *Sea $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere $\boldsymbol{Y}_1 = \boldsymbol{A}_1\boldsymbol{X}$, $\boldsymbol{Y}_2 = \boldsymbol{A}_2\boldsymbol{X}$ dos funciones lineales del vector aleatorio $\boldsymbol{X}$. La covarianza entre $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ es dada por*

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

Este resultado permite obtener una condición para la independencia entre dos formas lineales en variables aleatorias normales, estos es $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ serán independientes si y sólo si $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2^\top = \boldsymbol{0}$.

EJEMPLO 1.34. Considere $X_1, \dots, X_n$ una muestra aleatoria desde $\mathsf{N}(\mu, \sigma^2)$ y sea $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ y el vector de datos centrados $\boldsymbol{Z} = (Z_i)$, con $Z_i = X_i - \overline{X}$, $i = 1, \dots, n$. Note que, podemos escribir

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

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

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

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

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

La esfera p-dimesional es el conjunto de todos los puntos $\boldsymbol{x} \in \mathbb{R}^p$ tal que $\|\boldsymbol{x}\| = 1$. Una manera de definir la densidad de un vector aleatorio $\boldsymbol{U} \sim \mathsf{U}(\mathcal{S}^{(p)})$ se basa en la siguiente propiedad:

PROPIEDAD 1.36. Si $Z_1, \dots, Z_p \overset{\mathsf{ind}}{\sim} \mathsf{N}(0, 1)$, entonces $\boldsymbol{U} = (U_1, \dots, U_p)^\top \sim \mathsf{U}(\mathcal{S}^{(p)})$, donde

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

DEFINICIÓN 1.37. Un vector aleatorio $p \times 1$, $\boldsymbol{X}$ se dice que tiene simetría esférica si para cualquier $\boldsymbol{Q} \in \mathcal{O}_p$, sigue que

$$\boldsymbol{Q}\boldsymbol{X} \overset{\mathsf{d}}{=} \boldsymbol{X}.$$

Ejemplo 1.38. Sea $\boldsymbol{U} \sim \mathsf{U}(\mathcal{S}^{(p)})$, entonces es bastante obvio que $\boldsymbol{Q}\boldsymbol{U} \stackrel{\mathsf{d}}{=} \boldsymbol{U}$.

Definición 1.39. Un vector aleatorio p-dimensional tiene *distribución esférica* sólo si su función característica satisface

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

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

Ejemplo 1.40. Sea $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$, tenemos que

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

Resultado 1.41. *Sea $\psi(\boldsymbol{t}^\top \boldsymbol{t})$ la función característica del vector aleatorio $\boldsymbol{X}$. Entonces $\boldsymbol{X}$ tiene representación estocástica*

$$\boldsymbol{X} \stackrel{\mathsf{d}}{=} R\boldsymbol{U},$$

donde $\boldsymbol{U} \sim \mathsf{U}(\mathcal{S}^{(p)})$ y $R \sim F(\boldsymbol{X})$ son independientes.

Resultado 1.42. *Suponga que $\boldsymbol{X} \stackrel{\mathsf{d}}{=} R\boldsymbol{U} \sim \mathsf{S}_p(\psi)$ $(\mathsf{P}(\boldsymbol{X} = \boldsymbol{0}) = 0)$, entonces*

$$\|\boldsymbol{X}\| \stackrel{\mathsf{d}}{=} R, \qquad \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \stackrel{\mathsf{d}}{=} \boldsymbol{U}.$$

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

Definición 1.43. Un vector aleatorio $p \times 1$, $\boldsymbol{X}$ tiene *distribución de contornos elípticos* con parámetros $\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} \geq 0$ si

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

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

Observación 1.44. La función característica de $\boldsymbol{X} \sim \mathsf{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \psi)$ es de la forma

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

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

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

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

Definición 1.45. Se dice que el vector $\boldsymbol{X}$ tiene distribución de contornos elípticos si su función de densidad es de la forma

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

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

$$\int_0^\infty u^{p/2} g(u)\,\mathsf{d}u < \infty.$$

Observación 1.46. Asuma que $\boldsymbol{X} \sim \mathsf{EC}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \psi)$ con $\operatorname{rg}(\boldsymbol{\Sigma}) = k$. Entonces,

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

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

Ejemplo 1.47 (Distribución t de Student). La función generadora de densidad de un vector aleatorio con distribución t de Student asume la forma

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

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

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

donde $K_\nu(x)$ denota la función de Bessel modificada de segundo tipo.

Ejemplo 1.48 (Distribución Exponencial Potencia). Para la distribución Exponencial Potencia (Gómez et al., 1988), la función generadora de densidades es dada por

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

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

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

Note también que $R^{2\lambda} \sim \mathsf{Gama}(\frac{1}{2}, \frac{p}{2\lambda})$.

Definición 1.49. Sea $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma}$ matriz $p \times p$ definida positiva y H función de distribución de la variable aleatoria positiva W. Entonces, se dice que el vector aleatorio $\boldsymbol{X}$ sigue una *distribución de mezcla de escala normal* si su función de densidad asume la forma

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

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

Ejemplo 1.50 (Distribución Slash). Un vector aleatorio $\boldsymbol{X}$ tiene distribución Slash si su función de densidad es de la forma:

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

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

Observación 1.51. Un vector aleatorio $\boldsymbol{X} \sim \mathsf{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathsf{H})$ admite la representación

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

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

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

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 1.52. *Sea* $\boldsymbol{Z} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$ *y sea* $U = \boldsymbol{Z}^\top \boldsymbol{Z}$. *Entonces* $U \sim \chi^2_p$, *con función de densidad*

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

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

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

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

Definición 1.53 (Distribución chi-cuadrado no central). Si $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{I})$, entonces $U = \boldsymbol{Y}^\top \boldsymbol{Y}$ 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 1.54. *Sea* $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{I})$ *donde* $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p) \neq \boldsymbol{0}$ *y sea* $U = \boldsymbol{Y}^\top \boldsymbol{Y}$. *Entonces la función característica de* U *es dada por*

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

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

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

$$\varphi_U(t) = \prod_{j=1}^p \varphi_{Y_j^2}(t).$$

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

$$\begin{aligned}\varphi_{Y_j^2}(t) &= \int_{-\infty}^{\infty} \exp(ity_j^2)(2\pi)^{-1/2} \exp\{-\tfrac{1}{2}(y_j - \mu_j)^2\}\, \mathsf{d}y_j \\ &= \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\Big\{-\frac{(1-2it)}{2}\Big(y_j - \frac{\mu_j}{1-2it}\Big)^2\Big\}\, \mathsf{d}y_j \\ &\times \exp\Big\{\frac{\mu_j^2}{2}\Big(\frac{1}{1-2it}\Big) - \frac{\mu_j^2}{2}\Big\},\end{aligned}$$

de este modo,

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

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

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

□

Observación 1.55. Es interesante notar que la función característica de la variable $U = \boldsymbol{Y}^\top \boldsymbol{Y}$, puede ser escrita como

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

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

$$U|Z \sim \chi^2_{p+2z}, \qquad Z \sim \mathsf{Poisson}(\lambda), \tag{1.5}$$

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 (1.5) 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

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

mientras que la varianza de U puede ser calculada como

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

Resultado 1.56. *Si $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ donde $\boldsymbol{\Sigma}$ es matriz no singular. Entonces*

(a) $(\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_p$.
(b) $\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X} \sim \chi^2_p(\lambda)$, *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 $\boldsymbol{X}$ en variables aleatorias normales independientes. Considere $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$ con $\boldsymbol{B}$ no singular. Para probar (a), tome

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

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

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

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

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

y
$$\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}=\boldsymbol{Y}^\top\boldsymbol{B}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{Y}=\boldsymbol{Y}^\top\boldsymbol{Y},$$
que por definición tiene una distribución chi-cuadrado no central, con parámetro de no centralidad
$$\lambda=\frac{1}{2}(\boldsymbol{B}^{-1}\boldsymbol{\mu})^\top(\boldsymbol{B}^{-1}\boldsymbol{\mu})=\frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.$$
□

Definición 1.57 (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,
$$F=\frac{X_1/\nu_1}{X_2/\nu_2}\sim\mathsf{F}(\nu_1,\nu_2,\lambda),$$
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 1.58 (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,
$$G=\frac{U_1}{U_1+U_2}\sim\mathsf{Beta}(\nu_1,\nu_2,\lambda),$$
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 1.59 (Distribución t de Student no central). Si $Y\sim\mathsf{N}(\mu,\sigma^2)$ 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 $Z\sim\mathsf{N}(0,1)$, $U\sim\chi^2_\nu$, 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 parémetro de no centralidad $\delta=\lambda^2/2$. De este modo,
$$t^2_\nu(\lambda)\stackrel{\mathsf{d}}{=}\mathsf{F}(1,\nu,\lambda^2/2).$$

1.8. Distribución de formas cuadráticas

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

RESULTADO 1.60. *Si* $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{I})$ *y* $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ *es matriz simétrica. Entonces* $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X} \sim \chi^2_k(\theta)$ *sólo si* $\boldsymbol{A}$ *es idempotente, en cuyo caso los grados de libertad y el parámetro de no centralidad están dados por*

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

respectivamente.

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

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

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

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

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

$$\begin{aligned} \mathsf{E}\{\chi^2_k(\theta)\} = k + 2\theta &= \mathsf{E}(\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}) = \operatorname{tr}(\mathsf{E}(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{A}) \\ &= \operatorname{tr}((\boldsymbol{I} + \boldsymbol{\mu}\boldsymbol{\mu}^\top)\boldsymbol{A}) = k + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}, \end{aligned}$$

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

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

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

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

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

Tenemos que $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{\delta}, \boldsymbol{I})$ con $\boldsymbol{\delta} = \boldsymbol{P}^\top\boldsymbol{\mu}$, de modo que $Y_j^2 \sim \chi^2_1(\delta_j^2/2)$ con función característica

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

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

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

Como $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X} \sim \chi^2_k(\theta)$ tiene función característica

$$\varphi_{X^\top AX}(t) = (1 - 2it)^{-k/2} \exp\Big(\frac{2it\theta}{1 - 2it}\Big),$$

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

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

que es idempotente. Luego

$$\boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P} = (\boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P})(\boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P}) = \boldsymbol{P}^\top \boldsymbol{A}^2\boldsymbol{P} \quad \Longrightarrow \quad \boldsymbol{A}^2 = \boldsymbol{A}.$$

□

RESULTADO 1.61. *Si* $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *donde* $\boldsymbol{\Sigma}$ *es no singular y* $\boldsymbol{X}$*,* $\boldsymbol{\mu}$ *y* $\boldsymbol{\Sigma}$ *son particionados como*

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

donde $\boldsymbol{X}_1$*,* $\boldsymbol{\mu}_1$ *son* $k \times 1$ *y* $\boldsymbol{\Sigma}_{11}$ *es* $k \times k$*. Entonces*

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

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

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

Luego,

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

de donde sigue que $\boldsymbol{\Sigma}_{11} = \boldsymbol{B}_1\boldsymbol{B}_1^\top$. Ahora, sea $\boldsymbol{Z} = \boldsymbol{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$. De este modo,

$$\begin{pmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{pmatrix} \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{X}_2 - \boldsymbol{\mu}_2 \end{pmatrix}.$$

Entonces

$$\begin{aligned} U &= \boldsymbol{Z}^\top \boldsymbol{Z} - \boldsymbol{Z}^\top \boldsymbol{B}_1^\top (\boldsymbol{B}_1\boldsymbol{B}_1^\top)^{-1}\boldsymbol{B}_1\boldsymbol{Z} = \boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{B}_1^\top(\boldsymbol{B}_1\boldsymbol{B}_1^\top)^{-1}\boldsymbol{B}_1)\boldsymbol{Z} \\ &= \boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{Z}, \qquad \text{con } \boldsymbol{H}_1 = \boldsymbol{B}_1^\top(\boldsymbol{B}_1\boldsymbol{B}_1^\top)^{-1}\boldsymbol{B}_1. \end{aligned}$$

Note que $\boldsymbol{H}_1$ es simétrica e idempotente y por tanto también lo es $\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{H}_1$. De donde sigue que $U \sim \chi^2_\nu$, con $\nu = \operatorname{rg}(\boldsymbol{C}) = p - k$. □

El Resultado 1.60 se puede generalizar al caso que $\boldsymbol{X}$ tiene una matriz de covarianza arbitraria. Suponga que $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Una condición para que $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$ tenga una distribución chi-cuadrado es

$$\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{\Sigma}\boldsymbol{A},$$

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

RESULTADO 1.62. *Si* $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ *donde* $\boldsymbol{\Sigma}$ *tiene rango* k $(\leq p)$ *y si* $\boldsymbol{A}$ *es una inversa generalizada de* $\boldsymbol{\Sigma}$ $(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} = \boldsymbol{\Sigma})$*, entonces* $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X} \sim \chi^2_k$.

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

$$\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top = \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

Particionando $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ donde $\boldsymbol{Y}_1$ es un vector $k \times 1$ sigue que $\boldsymbol{Y}_1 \sim \mathsf{N}_k(\boldsymbol{0}, \boldsymbol{I})$ y $\boldsymbol{Y}_2 = \boldsymbol{0}$ con probabilidad 1. Es decir, tenemos que

$$\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{0})^\top, \quad \text{con probabilidad 1.}$$

Ahora, note que

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

Luego, con probabilidad uno,

$$\begin{aligned} \boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}^\top\boldsymbol{B}^{-\top}\boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{Y} &= (\boldsymbol{Y}_1^\top, \boldsymbol{0})\boldsymbol{B}^{-\top}\boldsymbol{A}\boldsymbol{B}^{-1}\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{0} \end{pmatrix} \\ &= (\boldsymbol{Y}_1^\top, \boldsymbol{0})\begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{B}^{-\top}\boldsymbol{A}\boldsymbol{B}^{-1}\begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{0} \end{pmatrix} \\ &= (\boldsymbol{Y}_1^\top, \boldsymbol{0})\begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{0} \end{pmatrix} = \boldsymbol{Y}_1^\top\boldsymbol{Y}_1 \sim \chi_k^2. \end{aligned}$$

□

RESULTADO 1.63. *Si* $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, donde* $\boldsymbol{\Sigma}$ *es no singular, y* $\boldsymbol{A}$ *es una matriz simétrica* $p \times p$*. Entonces* $\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X} \sim \chi_k^2(\lambda)$*, donde* $k = \operatorname{rg}(\boldsymbol{A})$*,* $\lambda = \boldsymbol{\mu}^\top\boldsymbol{A}\boldsymbol{\mu}/2$ *si y sólo si* $\boldsymbol{A}\boldsymbol{\Sigma}$ *es matriz idempotente.*

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

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

donde $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{I})$. Desde el Resultado 1.60 sigue que $\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}$ tiene distribución chi-cuadrado sólo si $\boldsymbol{B}^{-\top}\boldsymbol{A}\boldsymbol{B}^{-1}$ es idempotente. Esto es equivalente a mostrar que $\boldsymbol{A}\boldsymbol{\Sigma}$ es idempotente.

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

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

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

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

y por tanto es idempotente.

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

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

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

Ejemplo 1.64. Sea $X_1,\dots,X_n$ variables aleatorias IID $\mathsf{N}(\theta,\sigma^2)$, en este caso podemos definir $\boldsymbol{X}=(X_1,\dots,X_n)^\top$ tal que $\boldsymbol{X}\sim\mathsf{N}_n(\theta\boldsymbol{1}_n,\sigma^2\boldsymbol{I})$. Considere la forma cuadrática

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

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

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

que es idempotente. Además

$$\operatorname{rg}(\boldsymbol{A})=\operatorname{tr}\Big(\boldsymbol{I}_n-\frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top\Big)=n-1,$$

y

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

Finalmente,

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

Resultado 1.65. *Sea* $\boldsymbol{X}\sim\mathsf{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $Q_1=\boldsymbol{X}^\top\boldsymbol{A}_1\boldsymbol{X}$ *y* $Q_2=\boldsymbol{X}^\top\boldsymbol{A}_2\boldsymbol{X}$. *Entonces* Q_1 *y* Q_2 *son independientes si y sólo si* $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2=\boldsymbol{0}$.

Demostración. *(Sólo suficiencia)* Sea $\boldsymbol{Y}=\boldsymbol{B}^{-1}\boldsymbol{X}$ donde $\boldsymbol{\Sigma}=\boldsymbol{B}\boldsymbol{B}^\top$, entonces

$$Q_i=\boldsymbol{X}^\top\boldsymbol{A}_i\boldsymbol{X}=\boldsymbol{Y}^\top\boldsymbol{B}^{-\top}\boldsymbol{A}_i\boldsymbol{B}^{-1}\boldsymbol{Y}=\boldsymbol{Y}^\top\boldsymbol{G}_i\boldsymbol{Y},\quad i=1,2.$$

De este modo, $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2=\boldsymbol{0}$ implica que $\boldsymbol{G}_1\boldsymbol{G}_2=\boldsymbol{0}$. Tenemos que, existe una matriz ortogonal $\boldsymbol{P}$, tal que

$$\boldsymbol{P}^\top\boldsymbol{G}_i\boldsymbol{P}=\boldsymbol{M}_i,\qquad i=1,2,$$

es diagonal. Además,

$$\boldsymbol{M}_1\boldsymbol{M}_2=\boldsymbol{P}^\top\boldsymbol{G}_1\boldsymbol{P}\boldsymbol{P}^\top\boldsymbol{G}_2\boldsymbol{P}=\boldsymbol{P}^\top\boldsymbol{G}_1\boldsymbol{G}_2\boldsymbol{P}=\boldsymbol{0},$$

como $\boldsymbol{P}$ es matriz ortogonal, sigue que los elementos diagonales de $\boldsymbol{G}_1$ y $\boldsymbol{G}_2$ deben ocurrir en posiciones diferentes. Esto es, podemos escribir

$$\boldsymbol{G}_1=\begin{pmatrix}\boldsymbol{D}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\qquad\text{y}\qquad\boldsymbol{G}_2=\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}_2\end{pmatrix},$$

con $\boldsymbol{D}_1$ matriz diagonal cuya dimensión es dada por el rango de $\boldsymbol{A}_1$, digamos $r_1=\operatorname{rg}(\boldsymbol{A}_1)$. Sea $\boldsymbol{Z}=\boldsymbol{P}^\top\boldsymbol{Y}$, luego

$$Q_i=\boldsymbol{Y}^\top\boldsymbol{G}_i\boldsymbol{Y}=\boldsymbol{Z}^\top\boldsymbol{P}^\top\boldsymbol{G}_i\boldsymbol{P}\boldsymbol{Z}=\boldsymbol{Z}^\top\boldsymbol{M}_i\boldsymbol{Z}=\boldsymbol{Z}_i^\top\boldsymbol{D}_i\boldsymbol{Z}_i,$$

donde $\boldsymbol{Z}=(\boldsymbol{Z}_1^\top,\boldsymbol{Z}_2^\top)^\top$ ha sido particionado de acuerdo con $\boldsymbol{D}_1$ y $\boldsymbol{D}_2$.
Como

$$\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{Z}_1\\ \boldsymbol{Z}_2\end{pmatrix}=\boldsymbol{P}^\top\boldsymbol{Y}\sim\mathsf{N}_p(\boldsymbol{P}^\top\boldsymbol{B}\boldsymbol{\mu},\boldsymbol{I}),$$

tenemos que la independencia de Q_1 y Q_2 sigue desde la independencia entre $\boldsymbol{Z}_1$ y $\boldsymbol{Z}_2$. □

Resultado 1.66. *Sea* $\boldsymbol{X}\sim\mathsf{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $Q=\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X}$ *y* $\boldsymbol{U}=\boldsymbol{B}\boldsymbol{X}$. *Entonces* Q *y* $\boldsymbol{U}$ *son independientes si y sólo si* $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A}=\boldsymbol{0}$.

RESULTADO 1.67 (Teorema de Cochran). *Sea* $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *y* $\boldsymbol{A}_i$, $i = 1, \dots, k$ *matrices simétricas con* $\operatorname{rg}(\boldsymbol{A}_i) = r_i$ *y considere*

$$\boldsymbol{A} = \sum_{i=1}^{k} \boldsymbol{A}_i,$$

con $\operatorname{rg}(\boldsymbol{A}) = r$. *Si* $\boldsymbol{A\Sigma}$ *es idempotente y* $r = \sum_{i=1}^{k} r_i$, *entonces las formas cuadráticas* $Q_i = \boldsymbol{X}^\top \boldsymbol{A}_i \boldsymbol{X}$ *son variables aleatorias mutuamente independientes con distribución chi-cuadrado no central, es decir*

$$Q_i \sim \chi^2_{r_i}(\lambda_i),$$

donde $\lambda_i = \boldsymbol{\mu}^\top \boldsymbol{A} \boldsymbol{\mu}_i/2$, *para* $i = 1, \dots, k$.

DEMOSTRACIÓN. Suponga que $\boldsymbol{\Sigma} = \boldsymbol{I}$ y asuma $\boldsymbol{\mu} = \boldsymbol{0}$. Además, considere $\boldsymbol{A} = \boldsymbol{I}$ (sino, defina $\boldsymbol{A}_{k+1} = \boldsymbol{I} - \boldsymbol{A}$ con $\operatorname{rg}(\boldsymbol{A}_{k+1}) = r_{k+1} = p - r$, y la condición sobre el rango es satisfecha). Sea $\boldsymbol{P}_1$ matriz ortogonal que diagonaliza $\boldsymbol{A}_1$, de este modo

$$\boldsymbol{P}_1^\top \boldsymbol{A}_1 \boldsymbol{P}_1 = \boldsymbol{D}_1,$$

por conveniencia asuma que los elementos diagonales de $\boldsymbol{D}_1$ se encuentran en las primeras r_1 posiciones. Por tanto,

$$\boldsymbol{P}_1^\top (\boldsymbol{I} - \boldsymbol{A}_1) \boldsymbol{P}_1 = \boldsymbol{I} - \boldsymbol{D}_1 = \sum_{i=2}^{k} \boldsymbol{P}_1^\top \boldsymbol{A}_i \boldsymbol{P}_1,$$

las últimos $p - r_1$ elementos diagonales de $\boldsymbol{I} - \boldsymbol{D}_1$ son 1 y por tanto el rango de $\boldsymbol{I} - \boldsymbol{A}_1$ es al menos $p - r_1$.

Por otro lado, el rango de la suma de matrices $\sum_{i=2}^{k} \boldsymbol{P}_1^\top \boldsymbol{A}_i \boldsymbol{P}_1$ no puede ser mayor que la suma de los rangos de las matrices individuales, que por suposición es $p - r_1$. Por tanto $\operatorname{rg}(\boldsymbol{I} - \boldsymbol{A}_1) = p - r_1$. Esto implica que $\boldsymbol{P}_1^\top \boldsymbol{A}_1 \boldsymbol{P}_1$ tiene unos en las primeras r_1 posiciones (existen r_1 vectores linealmente independientes tal que $(\boldsymbol{I} - \boldsymbol{A}_1)\boldsymbol{m}_i = \boldsymbol{0}$, $i = 1, \dots, r_1$, esto es $\boldsymbol{A}_1 \boldsymbol{m}_i = \boldsymbol{m}_i$). Es decir, los valores propios no nulos de $\boldsymbol{A}_1$ son uno y de ahí que $\boldsymbol{A}_1$ es idempotente. Por este mismo razonamiento $\boldsymbol{A}_2, \dots, \boldsymbol{A}_k$ son idempotentes.

Es posible notar que las primeras r_1 filas y columnas de cada una de las matrices $\boldsymbol{P}_1^\top \boldsymbol{A}_i \boldsymbol{P}_1$ deben ser cero, pues estas matrices son simétricas y semidefinidas positivas, además si una matriz idempotente tiene un elemento diagonal cero toda la fila o columna asociada debe ser cero. De esto sigue que

$$\boldsymbol{I} - \boldsymbol{D}_1 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-r_1} \end{pmatrix} = \sum_{i=2}^{k} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_i \end{pmatrix}.$$

Sea P_2 matriz que diagonaliza la matriz idempotente $\boldsymbol{B}_2$ tal que $\boldsymbol{P}_2^\top \boldsymbol{B}_2 \boldsymbol{P}_2$ tiene elementos diagonales 1 en las primeras r_2 posiciones. Luego $\boldsymbol{P}_1 \boldsymbol{P}_2$ diagonaliza $\boldsymbol{A}_1$ y $\boldsymbol{A}_2$ tal que la matriz

$$\boldsymbol{P}_2^\top \boldsymbol{P}_1^\top (\boldsymbol{A}_1 + \boldsymbol{A}_2) \boldsymbol{P}_1 \boldsymbol{P}_2,$$

tiene unos en las primeras $r_1 + r_2$ posiciones y ceros en las posiciones restantes. Continuando con esta construcción se obtiene una matriz ortogonal $\boldsymbol{P}$ que diagonaliza $\boldsymbol{A}_i$, $i = 1, \dots, k$. En este caso $\boldsymbol{A}_i \boldsymbol{A}_j = \boldsymbol{0}$ para todo $i \neq j = 1, \dots, k$ que establece la independencia de las variables Q_i.

Se ha asumido que $\boldsymbol{X} \sim \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{I})$. Cuya función característica

$$\varphi(\boldsymbol{t}) = \mathsf{E}\{\exp(i\sum_{j=1}^{k} t_j q_j)\},$$

tiene exponente en la integral dado por

$$\sum_{j=1}^{k} it_j q_j - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x} = \sum_{j=1}^{k} it_j \boldsymbol{x}^\top \boldsymbol{A}_j \boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x}.$$

Considere la transformación $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Z}$ con $\boldsymbol{Z} = (\boldsymbol{Z}_1^\top, \boldsymbol{Z}_2^\top, \dots, \boldsymbol{Z}_k^\top)^\top$ donde cada $\boldsymbol{Z}_i$ tiene largo r_i, $i = 1, \dots, k$. De este modo el exponente resulta

$$-\frac{1}{2}\sum_{j=1}^{k}(1 - 2it_j)\boldsymbol{z}_j^\top \boldsymbol{z}_j,$$

como $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I})$ sigue que

$$\varphi(\boldsymbol{t}) = \prod_{j=1}^{k}(1 - 2it_j)^{-r_j},$$

que corresponde a la función característica de k variables aleatorias independientes chi-cuadrado con r_j grados de libertad. □