MAT-269: Conceptos preliminares

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Sea $\boldsymbol{x}$ vector aleatorio n-dimensional con densidad f. Entonces la esperanza de cualquier función $\boldsymbol{g}$ (de $\boldsymbol{x}$) es dada por:

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

siempre que la integral (n-dimensional) exista.

Considere $\boldsymbol{Z} = (z_{ij})$ una función matricial $m \times n$, entonces podemos definir la esperanza de una matriz aleatoria como:

$$\mathsf{E}(\boldsymbol{Z}(\boldsymbol{x})) = (\mathsf{E}(z_{ij})), \qquad z_{ij} = z_{ij}(\boldsymbol{x}).$$

Evidentemente

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

para $\boldsymbol{A} = (a_{ij})$ matriz de constantes.

Resultado 1:

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

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

Definición 1:

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 $m \times n$,

$$\operatorname{Cov}(\boldsymbol{x}, \boldsymbol{y}) = (\operatorname{Cov}(x_i, y_j)).$$

Observación:

Es fácil notar que:

$$\begin{aligned} \operatorname{Cov}(\boldsymbol{x}, \boldsymbol{y}) &= \operatorname{E}\{(\boldsymbol{x} - \operatorname{E}(\boldsymbol{x}))(\boldsymbol{y} - \operatorname{E}(\boldsymbol{y}))^\top\} \\ &= \operatorname{E}(\boldsymbol{x}\boldsymbol{y}^\top) - \operatorname{E}(\boldsymbol{x}) \operatorname{E}^\top(\boldsymbol{y}). \end{aligned}$$

Adicionalmente,

$$\operatorname{Cov}(\boldsymbol{x}) = \operatorname{Cov}(\boldsymbol{x}, \boldsymbol{x}) = \operatorname{E}(\boldsymbol{x}\boldsymbol{x}^\top) - \operatorname{E}(\boldsymbol{x}) \operatorname{E}^\top(\boldsymbol{x}),$$

es llamada matriz de dispersión.

Resultado 2:

Si $\boldsymbol{x}$ e $\boldsymbol{y}$ son vectores aleatorios m y n-dimensionales, respectivamente y $\boldsymbol{A} \in \mathbb{R}^{l \times m}$, $\boldsymbol{B} \in \mathbb{R}^{p \times n}$, entonces

$$\operatorname{Cov}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}\boldsymbol{y}) = \boldsymbol{A} \operatorname{Cov}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{B}^\top.$$

Observación:

Tenemos el siguiente caso particular,

$$\operatorname{Cov}(\boldsymbol{A}\boldsymbol{x}) = \operatorname{Cov}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A} \operatorname{Cov}(\boldsymbol{x}, \boldsymbol{x}) \boldsymbol{A}^\top = \boldsymbol{A} \operatorname{Cov}(\boldsymbol{x}) \boldsymbol{A}^\top.$$

Resultado 3:

Toda matriz de dispersión es simétrica y semidefinida positiva.

Demostración:

La simetría de la matriz de dispersión es obvia. Sea $\boldsymbol{z} = \boldsymbol{x} - \mathsf{E}(\boldsymbol{x})$, y considere la variable aleatoria $y = \boldsymbol{a}^\top \boldsymbol{z}$, para $\boldsymbol{a} \in \mathbb{R}^n$ un vector arbitrario. Entonces,

$$\begin{aligned}
\boldsymbol{a}^\top \mathsf{Cov}(\boldsymbol{x})\boldsymbol{a} &= \boldsymbol{a}^\top \mathsf{E}(\boldsymbol{x} - \mathsf{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathsf{E}(\boldsymbol{x}))^\top \boldsymbol{a} \\
&= \mathsf{E}(\boldsymbol{a}^\top (\boldsymbol{x} - \mathsf{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathsf{E}(\boldsymbol{x}))^\top \boldsymbol{a}) \\
&= \mathsf{E}(\boldsymbol{a}^\top \boldsymbol{z}\boldsymbol{z}^\top \boldsymbol{a}) = \mathsf{E}(y^2) \geq 0
\end{aligned}$$

y por tanto, $\mathsf{Cov}(\boldsymbol{x})$ es semidefinida positiva.

Suponga que $\mathsf{Cov}(\boldsymbol{x})$ es semidefinida positiva de rango r $(r \leq n)$. Luego

$$\mathsf{Cov}(\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{B}^\top,$$

donde $\boldsymbol{B} \in \mathbb{R}^{n\times r}$ de rango r.

Sea $\boldsymbol{y}$ vector aleatorio r-dimensional con $\mathsf{E}(\boldsymbol{y}) = \boldsymbol{0}$ y $\mathsf{Cov}(\boldsymbol{y}) = \boldsymbol{I}$. Haciendo $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}$, sigue que $\mathsf{E}(\boldsymbol{x}) = \boldsymbol{0}$ y

$$\mathsf{Cov}(\boldsymbol{x}) = \mathsf{Cov}(\boldsymbol{B}\boldsymbol{y}) = \boldsymbol{B}\,\mathsf{Cov}(\boldsymbol{y})\boldsymbol{B}^\top = \boldsymbol{B}\boldsymbol{B}^\top.$$

Es decir, corresponde a una matriz de covarianza.

Resultado 4:

Sea $\boldsymbol{x}$ vector aleatorio n-dimensional y considere la transformación lineal

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},$$

donde $\boldsymbol{A}$ es una matriz de constantes $m \times n$ y $\boldsymbol{b}$ es vector de constantes $m \times 1$. Entonces

$$\mathsf{E}(\boldsymbol{y}) = \boldsymbol{A}\,\mathsf{E}(\boldsymbol{x}) + \boldsymbol{b}, \qquad \mathsf{Cov}(\boldsymbol{y}) = \boldsymbol{A}\,\mathsf{Cov}(\boldsymbol{x})\boldsymbol{A}^\top.$$

Ejemplo (estandarización):

Sea $\boldsymbol{x}$ vector aleatorio n-dimensional con media $\mathsf{E}(\boldsymbol{x}) = \boldsymbol{\mu}$ y matriz de dispersión $\mathsf{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$, tal que podemos escribir

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

donde $\boldsymbol{U}$ es matriz ortogonal y $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda})$. Considere,

$$\boldsymbol{z} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^\top(\boldsymbol{x} - \boldsymbol{\mu})$$

de este modo, obtenemos que

$$\mathsf{E}(\boldsymbol{z}) = \boldsymbol{0} \qquad \text{y} \qquad \mathsf{Cov}(\boldsymbol{z}) = \boldsymbol{I}.$$

Ejemplo (Whitening):

Para $\boldsymbol{x} = (x_1, x_2, \dots, x_p)^\top$ con $\mathsf{E}(\boldsymbol{x}) = \boldsymbol{\mu}$ y $\mathsf{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$. La transformación,

$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x},$$

tal que $\mathsf{Cov}(\boldsymbol{z}) = \boldsymbol{I}$ es conocida como whitening.

Evidentemente, debemos tener que

$$\mathsf{Cov}(\boldsymbol{z}) = \boldsymbol{W} \mathsf{Cov}(\boldsymbol{x}) \boldsymbol{W}^\top = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}^\top = \boldsymbol{I},$$

es decir,

$$\boldsymbol{W}(\boldsymbol{\Sigma}\boldsymbol{W}^\top\boldsymbol{W}) = \boldsymbol{W}, \qquad \Longrightarrow \qquad \boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}.$$

Observación:

Existen varios métodos tradicionales para obtener $\boldsymbol{W}$ (Kessy et al., 2018)[1]. Por ejemplo:

- PCA whitening: basado en $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, obtenemos
$$\boldsymbol{W}_{\mathsf{PCA}} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^\top.$$
- Cholesky whitening: considere $\boldsymbol{\Sigma}^{-1} = \boldsymbol{L}\boldsymbol{L}^\top$, donde $\boldsymbol{L}$ es triangular inferior. Entonces,
$$\boldsymbol{W}_{\mathsf{Chol}} = \boldsymbol{L}^\top.$$

[1] The American Statistician **72**, 309-314.

Definición 2:

Sea $\boldsymbol{x} = (x_1, \dots, x_p)^\top$ vector aleatorio con media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$. Se define la matriz de correlación como $\boldsymbol{R} = (\rho_{ij})$, donde

$$\rho_{ij} = \frac{\mathsf{Cov}(x_i, x_j)}{\{\mathsf{var}(x_i)\,\mathsf{var}(x_j)\}^{1/2}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \qquad i, j = 1, \dots, p.$$

Observación:

Considere $\boldsymbol{D} = \operatorname{diag}(\sigma_{11}, \dots, \sigma_{pp}) = \operatorname{diag}(\boldsymbol{\Sigma})$, entonces podemos escribir

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

(y análogamente $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}$).

Definición 3:

Si el vector aleatorio $\boldsymbol{x}$ tiene función característica dada por

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

decimos que sigue una distribución normal multivariada y anotamos $\boldsymbol{x} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Resultado 5:

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

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

Resultado 6:

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

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

Observación:

Sea

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{z}, \qquad \boldsymbol{z} \sim \mathsf{N}_p(\mathbf{0}, \boldsymbol{I}),$$

donde $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\top$ con $\boldsymbol{B}$ matriz de rango completo. Por el Resultado 5, sigue que

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

Además,

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

Considere la siguiente partición:

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{1}$$

donde $\boldsymbol{x}_1$ y $\boldsymbol{\mu}_1$ son vectores $k \times 1$ y $\boldsymbol{\Sigma}_{11}$ es $k \times k$.

Resultado 7:

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.

Resultado 8:

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). Entonces los vectores $\boldsymbol{x}_1$ y $\boldsymbol{x}_2$ son independientes sólo si $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Resultado 9:

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

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

Ejemplo:

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

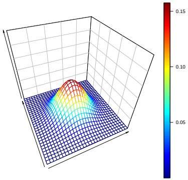
$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad -1 < \rho < 1.$$

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

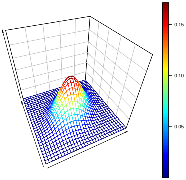
$$f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}(x_1^2 + x_2^2 - 2\rho x_1 x_2)\Big\}.$$

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

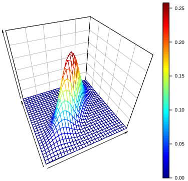
Distribución normal multivariada



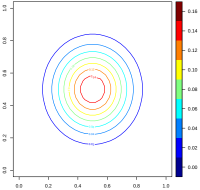
(a) $\rho = 0.0$



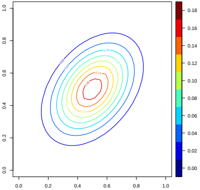
(b) $\rho = 0.4$



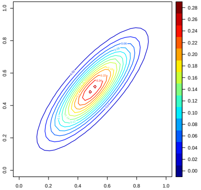
(c) $\rho = 0.8$



(d) contornos



(e) contornos



(f) contornos

Observación:

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

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

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

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{z}^\top \boldsymbol{z} = \sum_{i=1}^{p} z_i^2,$$

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[2] de $\boldsymbol{x}$ a $\boldsymbol{\mu}$.

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

Resultado 10:

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

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