

MGE-201: Conceptos preliminares

Felipe Osorio

felipe.osorio@uv.cl

Conceptos preliminares

Un vector aleatorio k -dimensional \mathbf{X} es una función desde el espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ a \mathbb{R}^k , esto es

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

Definición 1:

La función de distribución de \mathbf{X} es una función $F : \mathbb{R}^k \rightarrow [0, 1]$, tal que

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

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

La notación anterior debe ser entendida como:

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

que corresponde a la probabilidad del evento $\bigcap_{r=1}^k \{X_r \leq x_r\}$.

Conceptos preliminares

Propiedades:

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

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

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

en este caso decimos que \mathbf{X} es un vector aleatorio continuo con función de densidad f . Además,

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

Conceptos preliminares

Considere el vector aleatorio k -dimensional \mathbf{X} particionado como $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ donde \mathbf{X}_1 y \mathbf{X}_2 son vectores $k_1 \times 1$ y $k_2 \times 1$, respectivamente, con $k = k_1 + k_2$.

En este caso $\mathbf{X}_i \sim F_i$, $i = 1, 2$, y \mathbf{X}_1 , \mathbf{X}_2 son llamadas **marginales** de \mathbf{X} .

Note que

$$F_1(\mathbf{s}) = F(\mathbf{s}, +\infty), \quad F_2(\mathbf{t}) = F(+\infty, \mathbf{t}), \quad \forall \mathbf{s} \in \mathbb{R}^{k_1}, \mathbf{t} \in \mathbb{R}^{k_2},$$

o bien

$$f_1(\mathbf{s}) = \int_{\mathbb{R}^{k_2}} f(\mathbf{s}, \mathbf{u}) d\mathbf{u}, \quad f_2(\mathbf{t}) = \int_{\mathbb{R}^{k_1}} f(\mathbf{u}, \mathbf{t}) d\mathbf{u},$$

para $\mathbf{s} \in \mathbb{R}^{k_1}$ y $\mathbf{t} \in \mathbb{R}^{k_2}$.

Conceptos preliminares

Si \mathbf{X} es absolutamente continuo y $f_1(\mathbf{x}_1) > 0$, entonces la **densidad condicional** de \mathbf{X}_2 dado $\mathbf{X}_1 = \mathbf{x}_1$ es dada por

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

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

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

tenemos además que

$$f_{X_2|X_1=x_1}(\mathbf{u}) = \frac{f_X(\mathbf{x}_1, \mathbf{u})}{\int_{\mathbb{R}^{k_2}} f_X(\mathbf{x}_1, \mathbf{t}) dt}.$$

Conceptos preliminares

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

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

siempre que la integral (k -dimensional) exista.

Más generalmente, sea $\mathbf{Z} = (Z_{ij})$ una matriz aleatoria $m \times n$, entonces

$$\mathbb{E}(\mathbf{Z}) = (\mathbb{E}(Z_{ij})).$$

En particular, para $\mathbf{A} = (a_{ij})$ una matriz de constantes, entonces

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

Conceptos preliminares

Resultado 1:

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

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

Demostración:

Sea $\mathbf{Y} = \mathbf{A}\mathbf{Z}\mathbf{B} + \mathbf{C}$, entonces

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

de este modo

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

Conceptos preliminares

Definición 2:

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

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

Además, es fácil mostrar que

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

En particular,¹

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

¹Cov(\mathbf{X}) también es conocida como matriz de dispersión.

Conceptos preliminares

Resultado 2:

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

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

Demostración:

Sean $\mathbf{U} = \mathbf{AX}$ y $\mathbf{V} = \mathbf{BY}$, entonces

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

Observación:

En particular,

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

Conceptos preliminares

Resultado 3:

Toda matriz de dispersión es simétrica y semidefinida positiva.

Demostración:

La simetría es directo. Para mostrar que $\text{Cov}(\mathbf{X})$ es semidefinida positiva, sea $\mathbf{Z} = \mathbf{X} - \mathbb{E}(\mathbf{X})$, y considere $Y = \mathbf{a}^\top \mathbf{Z}$, para $\mathbf{a} \in \mathbb{R}^n$. Entonces,

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

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

Conceptos preliminares

Suponga que $\text{Cov}(\mathbf{X})$ es semidefinida positiva de rango r ($r \leq n$), de ahí que $\text{Cov}(\mathbf{X}) = \mathbf{B}\mathbf{B}^\top$ donde $\mathbf{B} \in \mathbb{R}^{n \times r}$ de rango r .

Sea \mathbf{Y} vector aleatorio r -dimensional con $E(\mathbf{Y}) = \mathbf{0}$ y $\text{Cov}(\mathbf{Y}) = \mathbf{I}$. Haciendo $\mathbf{X} = \mathbf{B}\mathbf{Y}$, sigue que $E(\mathbf{X}) = \mathbf{0}$ y

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

Es decir, corresponde a una matriz de covarianza.

Conceptos preliminares

Resultado 4:

Sea \mathbf{X} vector aleatorio n -dimensional y considere la transformación lineal

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b},$$

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

Entonces

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

Conceptos preliminares

Definición 3:

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

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

Observación:

En efecto, podemos escribir,²

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

donde $\mathbf{D} = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$.

²O análogamente, $\boldsymbol{\Sigma} = \mathbf{D}^{1/2} \mathbf{R} \mathbf{D}^{1/2}$.

Conceptos preliminares

Resultado 5:

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

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

Demostración:

Tenemos

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

Conceptos preliminares

Ejemplo:

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

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

De este modo, tenemos

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

Llamaremos a $\mathbf{C} = \mathbf{I} - \frac{1}{n}\mathbf{J}_n$ la *matriz de centrado*.

Conceptos preliminares

Suponga que X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 . Sigue que,

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

pues $\text{Cov}(X_i, X_j) = 0$ ($i \neq j$). Por tanto, podemos usar el [Resultado 5](#) para calcular la esperanza de la variable aleatoria,

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

obteniendo

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

Es fácil verificar que

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

$$\mathbf{C} \mathbf{1} = \left(\mathbf{I} - \frac{1}{n} \mathbf{1} \mathbf{1}^\top\right) \mathbf{1} = \mathbf{1} - \frac{1}{n} \mathbf{1} \mathbf{1}^\top \mathbf{1} = \mathbf{1} - \mathbf{1} = \mathbf{0},$$

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

Conceptos preliminares

Definición 4:

La función generadora de momentos de \mathbf{X} es dada por:

$$M_X(\mathbf{t}) = \mathbb{E}\{\exp(\mathbf{t}^\top \mathbf{X})\}, \quad \|\mathbf{t}\| < h.$$

Observación:

$M_X(\mathbf{t})$ permite un método bastante operativo para el cálculo del k -ésimo momento de un vector aleatorio \mathbf{X} . En efecto,

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

Conceptos preliminares

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

$$F_Z(\mathbf{x}, \mathbf{y}) = G_X(\mathbf{x})H_Y(\mathbf{y}),$$

donde $F_Z(\mathbf{z})$, $G_X(\mathbf{x})$ y $H_Y(\mathbf{y})$ son las funciones de distribución de \mathbf{Z} , \mathbf{X} e \mathbf{Y} , respectivamente.

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

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

Desde esto, sigue que:

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

Conceptos preliminares

Resultado 6 (Transformación de vectores aleatorios):

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

$$f_Y(\mathbf{y}) = |\mathrm{D}\mathbf{g}^{-1}(\mathbf{y})|_+ f_X(\mathbf{g}^{-1}(\mathbf{y})),$$

donde

$$\mathrm{D}\mathbf{g}^{-1}(\mathbf{y}) = \frac{\partial \mathbf{g}^{-1}(\mathbf{y})}{\partial \mathbf{y}^\top} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}^\top},$$

denota la matriz Jacobiana asociada a la transformación $\mathbf{g}^{-1}(\cdot)$.