1. La normal multidimensional

Estadística II, 20-21

Pablo Fernández Gallardo

Vectores aleatorios

Vector aleatorio n dimensional:

$$\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Su distribución (en el caso continuo) viene dada por una función de densidad conjunta $f_{\mathbb{X}}(x_1, \ldots, x_n)$tal que

$$f_{\mathbb{X}}(\mathbf{x}) \geq 0 \quad \text{y} \qquad \int_{\mathbb{R}^n} f_{\mathbb{X}}(\mathbf{x})\, d\mathbf{x} = 1.$$

(La independencia de las coordenadas significa que la función de densidad se factoriza: $f_{\mathbb{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$).

Vector de medias

Notación: si $\mathbb{M} = (X_{i,j})_{i,j}$ es una *matriz* de dimensiones $n \times m$ cuyas componentes son variables aleatorias, escribiremos $\mathbf{E}(\mathbb{M})$ para referirnos la matriz $(\mathbf{E}(X_{i,j}))_{i,j}$ de medias de esas variables.

El vector de medias asociado a $\mathbb{X}$ es

$$\mathbf{E}(\mathbb{X}) = \begin{pmatrix} \mathbf{E}(X_1) \\ \vdots \\ \mathbf{E}(X_n) \end{pmatrix}.$$

Si A, B son matrices $n \times n$, $\mathbb{X}, \mathbb{Y}$ vectores aleatorios de dimensión n, y $\mathbf{b} \in \mathbb{R}^n$,

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

(Atención: $\mathbf{E}(\mathbb{X}^\intercal A\,\mathbb{X}) \neq \mathbf{E}(\mathbb{X}^\intercal)\,A\,\mathbf{E}(\mathbb{X})$).

Matriz de varianzas/covarianzas

La matriz de covarianzas de $\mathbb{X}$ es

$$\mathbf{cov}(\mathbb{X}) = \begin{pmatrix} \mathbf{V}(X_1) & \mathrm{cov}(X_1, X_2) & \cdots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathbf{V}(X_2) & \cdots & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \cdots & \mathbf{V}(X_n) \end{pmatrix}$$

Recuérdese que

$$\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2,$$
$$\mathrm{cov}(X, Y) = \mathbf{E}\big[(X - \mathbf{E}(X)) \cdot (Y - \mathbf{E}(Y))\big] = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y).$$

1) Se tiene que

$$\mathbf{cov}(\mathbb{X}) = \mathbf{E}\big((\mathbb{X} - \mathbf{E}(\mathbb{X})) \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))^\intercal\big)$$

(Atención: $(\mathbb{X} - \mathbf{E}(\mathbb{X})) \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))^\intercal$ es una matriz $n \times n$; mientras que $(\mathbb{X} - \mathbf{E}(\mathbb{X}))^\intercal \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))$ es una variable aleatoria).

2) Bajo cambios lineales,

$$\mathbf{cov}(A\mathbb{X} + \mathbf{b}) = A \cdot \mathbf{cov}(\mathbb{X}) \cdot A^\intercal.$$

3) Una matriz de covarianzas es simétrica y (semi)definida positiva.

Prueba: calcula $\mathbf{V}(\mathbf{a}^\intercal \mathbb{X})$.

Matriz de correlaciones

La matriz de correlaciones de $\mathbb{X}$ es

$$\boldsymbol{\rho}(\mathbb{X}) = \begin{pmatrix} 1 & \rho(X_1, X_2) & \cdots & \rho(X_1, X_n) \\ \rho(X_2, X_1) & 1 & \cdots & \rho(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_n, X_1) & \rho(X_n, X_2) & \cdots & 1 \end{pmatrix}$$

Recuérdese que

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathbf{V}(X)\mathbf{V}(Y)}} \in [-1, 1].$$

- Se tiene que $\mathbf{cov}(\mathbb{X}) = \sqrt{D(\mathbb{X})}\,\boldsymbol{\rho}(\mathbb{X})\sqrt{D(\mathbb{X})}$, donde $\sqrt{D(\mathbb{X})}$ es la matriz diagonal con las desviaciones típicas.

- ¿Cómo cambia $\boldsymbol{\rho}(\mathbb{X})$ bajo transformaciones lineales?

La normal multidimensional

Dadas

- un vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\mathsf{T}$;
- y una matriz $V = (v_{ij})$ de dimensiones $n \times n$ simétrica y definida positiva,

decimos que $\mathbb{X}$ sigue una distribución normal multidimensional (de dimensión n) con parámetros $\boldsymbol{\mu}$ y V, lo que denotaremos por

$$\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V),$$

si, para todo $\mathbf{x} \in \mathbb{R}^n$, su función de densidad viene dada por

$$f_{\mathbb{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(V)}}\; e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\mathsf{T}\, V^{-1}\,(\mathbf{x}-\boldsymbol{\mu})}.$$

En esta expresión, $\mathbf{x}$ es un vector columna.

Como V es simétrica y definida positiva, la podremos escribir como

$$V = UU^{\intercal},$$

para cierta matriz U no singular. Obsérvese que $\det(V) = \det(U)^2$. Como

$$V^{-1} = (U^{-1})^{\intercal} U^{-1},$$

U^{-1} es una raíz cuadrada de V^{-1}.

Podemos reescribir la densidad $f_{\mathbb{X}}$ como

$$f_{\mathbb{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(U)|} \, e^{-\frac{1}{2} \|U^{-1}(\mathbf{x}-\mathbf{m})\|^2}.$$

Por ejemplo, si orto-diagonalizamos V,

$$V = O\Lambda O^{\intercal} = \underbrace{O\,\Lambda^{1/2} O^{\intercal}}_{=U}\ \underbrace{O\,\Lambda^{1/2} O^{\intercal}}_{=U^{\intercal}}$$

y U^{-1} sería

$$U^{-1} = O\,\Lambda^{-1/2} O^{\intercal}.$$

Algunas observaciones:

- El caso $\mathcal{N}(\mathbf{0}, I)$ es la normal estándar.
- Si $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$, entonces
  - $\mathbf{E}(\mathbb{X}) = \boldsymbol{\mu}$ y $\mathbf{cov}(\mathbb{X}) = V$,
  - y cada X_i es una normal de media μ_i y varianza v_{ii}.
- Las coordenadas $X_1, \dots, X_n$ son independientes si y solo si V es diagonal.
- Tipificación:

$$\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V) \qquad \Longleftrightarrow \qquad \mathbb{X} = \boldsymbol{\mu} + U\,\mathbb{Y}, \quad \text{con } \mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I).$$

Es decir,

$$U^{-1}(\mathbb{X} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, I).$$

Si $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$, la variable aleatoria

$$(\mathbb{X} - \boldsymbol{\mu})^\intercal V^{-1}(\mathbb{X} - \boldsymbol{\mu})$$

se distribuye como una χ^2 con n grados de libertad.

Recuérdese que una χ^2 con n grados de libertad es una suma de n cuadrados de normales estándar. Basta observar que $(\mathbb{X} - \boldsymbol{\mu})^\intercal V^{-1}(\mathbb{X} - \boldsymbol{\mu}) = \mathbb{Y}^\mathsf{T}\mathbb{Y}$, con $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$.

Combinaciones lineales de coordenadas

Sea $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$ una normal de dimensión n.

- Para un vector $\mathbf{a}$, la variable aleatoria

$$\mathbf{a}^\mathsf{T}\mathbb{X} = \sum_{j=1}^{n} a_j\, X_j$$

  se distribuye como una $\mathcal{N}(\mathbf{a}^\mathsf{T}\boldsymbol{\mu}, \mathbf{a}^\mathsf{T} V \mathbf{a})$.
- Para una matriz A y un vector $\mathbf{b}$, el vector aleatorio

$$A\mathbb{X} + \mathbf{b}$$

  se distribuye como una $\mathcal{N}(A\boldsymbol{\mu}, AVA^\mathsf{T})$.

Reducción de dimensión

Sea $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$ una normal de dimensión n.

Seleccionamos k índices, $J = \{j_1, \ldots, j_k\}$ y definimos $\mathbb{X}_J$, $\boldsymbol{\mu}_J$ y V_J quedándonos con las entradas correspondientes de $\mathbb{X}_J$, $\boldsymbol{\mu}$ y V.

Entonces $\mathbb{X}_J \sim \mathcal{N}(\boldsymbol{\mu}_J, V_J)$.

Condicionando

Sea $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$ una normal de dimensión n.

Digamos que partimos en

$$\mathbb{X} = \left(\begin{array}{c} X_1 \\ \vdots \\ X_p \\ \hline X_{p+1} \\ \vdots \\ X_n \end{array}\right) = \left(\begin{array}{c} \mathbb{X}_1 \\ \hline \mathbb{X}_2 \end{array}\right)$$

Partimos, análogamente,

$$\boldsymbol{\mu} = \left(\begin{array}{c} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{array}\right) \qquad \text{y} \qquad V = \left(\begin{array}{|c|c|} V_{1,1} & V_{1,2} \\ \hline V_{2,1} & V_{2,2} \end{array}\right)$$

Entonces

- Los vectores $\mathbb{X}_1$ y $\mathbb{X}_2$ son independientes si $V_{1,2} = V_{2,1}^{\intercal}$ tiene todas sus entradas nulas.
- El vector $\mathbb{X}_1$, condicionado a que $\mathbb{X}_2 = \mathbf{a}$, se distribuye como una normal (de dimensión p) con parámetros

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + V_{1,2}\, V_{2,2}^{-1}(\mathbf{a} - \boldsymbol{\mu}_2)$$
$$\tilde{V} = V_{1,1} - V_{1,2} V_{2,2}^{-1} V_{2,1}.$$

Formas lineales y cuadráticas, y normalidad

Sea $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$.

Para ciertos cálculos del curso, interesará conocer la distribución de combinaciones lineales y cuadráticas de las coordenadas de $\mathbb{X}$:

$$\begin{array}{rl} \text{vectores aleatorios del tipo} & A\mathbb{X}, \\ \text{o variables aleatorias del tipo} & \mathbb{X}^{\mathsf{T}} A\, \mathbb{X}, \end{array}$$

y sus posibles relaciones de dependencia.

Recordamos que una variable Z se distribuye como una χ^2 con n grados de libertad si

$$Z = Z_1^2 + \cdots + Z_n^2,$$

donde $Z_1, \ldots, Z_n$ son normales estándar independientes. Se tiene que $\mathbf{E}(Z) = n$ y $\mathbf{V}(Z) = 2n$.

Caso lineal (ya visto):

- Si $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$ es una normal de dimensión n, y A es una matriz $n \times n$, entonces
$$A\mathbb{X} \sim \mathcal{N}(A\boldsymbol{\mu}, AVA^{\mathsf{T}}).$$

(En realidad, A podría ser de dimensiones $p \times n$, con $p < n$).

En el caso de las formas cuadráticas, tenemos que

- Si $\mathbb{X} \sim \mathcal{N}(\mathbf{0}, I)$ es una normal de dimensión n, entonces

$$\mathbb{X}^{\mathsf{T}}\, \mathbb{X} \sim \chi_n^2.$$

También hemos visto que

- si $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, V)$ es una normal de dimensión n, entonces

$$(\mathbb{X} - \boldsymbol{\mu})^{\mathsf{T}}\, V^{-1}(\mathbb{X} - \boldsymbol{\mu}) \sim \chi_n^2.$$

Teorema 1

Si $\mathbb{X} \sim \mathcal{N}(\mathbf{0}, I)$ es una normal de dimensión n, y B es una matriz $n \times n$ simétrica e idempotente, entonces

$$\mathbb{X}^\intercal B \, \mathbb{X} \sim \chi^2_{\mathrm{traza}(B)}.$$

Corolario 2

Si $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$ es una normal de dimensión n, y B es una matriz $n \times n$ simétrica e idempotente, y se tiene que $\boldsymbol{\mu}^\intercal B \boldsymbol{\mu} = 0$, entonces

$$\frac{1}{\sigma^2} \, \mathbb{X}^\intercal B \, \mathbb{X} \sim \chi^2_{\mathrm{traza}(B)}.$$

Sobre independencia

Teorema 3

Sea $\mathbb{X} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$ una normal de dimensión n. Sean

- A una matriz $p \times n$, con $p \leq n$;
- B y C matrices $n \times n$ simétricas e idempotentes.

Entonces,

- si $AB = \mathbf{0}$, entonces $A\mathbb{X}$ y $\mathbb{X}^\intercal B\mathbb{X}$ son independientes;
- si $BC = \mathbf{0}$, entonces $\mathbb{X}^\intercal B\mathbb{X}$ y $\mathbb{X}^\intercal C\mathbb{X}$ son independientes.