MAT-269: Sesión 7
Estimación bajo restricciones sobre μ y Σ

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Estimadores ML restringidos

Suponga $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ vectores aleatorios IID desde $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere:[1]

1. $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ conocido. Entonces,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{\mu}_0)^\top.$$

2. $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ conocido. De este modo,

$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}.$$

3. $\boldsymbol{\mu} = \gamma \boldsymbol{a}$, $\gamma \in \mathbb{R}$ con $\boldsymbol{a} \in \mathbb{R}^p$ conocido. Luego,

$$\widehat{\gamma}_\Sigma = \frac{\boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \overline{\boldsymbol{x}}}{\boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}}.$$

   Para $\boldsymbol{\Sigma}$ desconocido, tenemos:

$$\widehat{\gamma} = \frac{\boldsymbol{a}^\top \boldsymbol{S}^{-1} \overline{\boldsymbol{x}}}{\boldsymbol{a}^\top \boldsymbol{S}^{-1} \boldsymbol{a}}.$$

[1] Ud. lo deberá resolver como parte de la Tarea 1 (Entrega: 4 Mayo).

4. $\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{a}$, $\boldsymbol{A} \in \mathbb{R}^{q\times p}$, $\boldsymbol{a} \in \mathbb{R}^{q}$ matrices conocidas. Luego
$$\widehat{\boldsymbol{\mu}}_{\Sigma} = \overline{\boldsymbol{x}} - \boldsymbol{\Sigma}\boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top})^{-1}(\boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{a}),$$
para $\boldsymbol{\Sigma}$ desconocido
$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}} - \boldsymbol{S}\boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\top})^{-1}(\boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{a}).$$
5. $\boldsymbol{\Sigma} = \phi\boldsymbol{V}$ con $\boldsymbol{V} > 0$ conocida y $\phi > 0$. Por tanto,
$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}, \qquad \widehat{\phi} = \frac{1}{p}\operatorname{tr}(\boldsymbol{V}^{-1}\boldsymbol{S}).$$

Suponga que

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \qquad \text{con} \qquad \boldsymbol{\Sigma}_{12} = \mathbf{0} = \boldsymbol{\Sigma}_{21}^\top,$$

y $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$. De este modo, tenemos $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$, así la función de log-verosimilitud adopta la forma:

$$\ell(\boldsymbol{\theta}) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}).$$

Note que el estimador ML para $\boldsymbol{\mu}$ no depende de $\boldsymbol{\Sigma}$, luego $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$. Ahora, usando que

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{vmatrix} = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22}|, \qquad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix}.$$

Sigue que la parte relevante de $\ell(\boldsymbol{\theta})$ es dada por:

$$\begin{aligned}\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log|\boldsymbol{\Sigma}_{11}| - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)\\ - \frac{n}{2}\log|\boldsymbol{\Sigma}_{22}| - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\end{aligned}$$

que puede ser escrita como:

$$\ell(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}_1) + \ell_2(\boldsymbol{\theta}_2),$$

con $\boldsymbol{\theta}_j = (\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{jj})$, para $j = 1, 2$, y

$$\ell_j(\boldsymbol{\theta}_j) = -\frac{n}{2}\log|\boldsymbol{\Sigma}_{jj}| - \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{Q}_j(\boldsymbol{\mu}_j),$$

donde

$$\boldsymbol{Q}_j(\boldsymbol{\mu}_j) = \sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_j)(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^\top.$$

Matriz de covarianza diagonal

De ahí que, usando el Resultado 2 de la Sesión 6, obtenemos el estimador ML para $\mathbf{\Sigma}$:

$$\widehat{\mathbf{\Sigma}} = \begin{pmatrix} \widehat{\mathbf{\Sigma}}_{11} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Sigma}}_{22} \end{pmatrix},$$

con

$$\widehat{\mathbf{\Sigma}}_{jj} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_j)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_j)^\top,$$

para $j = 1, 2$, donde $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}_1^\top, \overline{\boldsymbol{x}}_2^\top)^\top$.

Muestras con parámetros 'enlazados'

Suponga que la matriz de datos $\boldsymbol{X}$ es particionada como:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_k \end{pmatrix},$$

donde las filas de $\boldsymbol{X}_i \in \mathbb{R}^{n_i \times p}$ son IID $\mathsf{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$, para $i = 1, \dots, k$.

Las restricciones más comunes son:

(a) $\boldsymbol{\Sigma}_{11} = \cdots = \boldsymbol{\Sigma}_{kk}$ (digamos, $= \boldsymbol{\Sigma}$).

(b) $\boldsymbol{\Sigma}_{11} = \cdots = \boldsymbol{\Sigma}_{kk}$ y $\boldsymbol{\mu}_1 = \cdots \boldsymbol{\mu}_k$.

Para el caso en (a) note que podemos escribir:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{k} \ell_i(\boldsymbol{\theta}),$$

donde

$$\ell_i(\boldsymbol{\theta}) = -\frac{n_i}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{Q}_i + n_i(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}_i)(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}_i)^\top\},$$

Sea

$$\boldsymbol{S}_i = \frac{1}{n_i}\boldsymbol{Q}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_i)^\top,$$

para $i = 1, \dots, k$. Es decir,

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{k}\Big\{n_i \log|\boldsymbol{\Sigma}| + n_i \operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{S}_i + (\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}_i)(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}_i)^\top)\Big\}.$$

Como no existe restricciones sobre $\boldsymbol{\mu}$ sigue que el MLE de $\boldsymbol{\mu}$ es $\overline{\boldsymbol{x}}_i$ $(i = 1, \dots, k)$. Además, considere

$$\boldsymbol{W} = \sum_{i=1}^{k} n_i \boldsymbol{S}_i, \qquad n = \sum_{i=1}^{k} n_i.$$

De este modo, la log-verosimilitud perfilada, es dada por:

$$\ell_*(\boldsymbol{\Sigma}) = \ell(\widehat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}) = -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{W}.$$

De ahí que el MLE para $\boldsymbol{\Sigma}$ adopta la forma:[2]

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \boldsymbol{W} = \frac{1}{n} \sum_{i=1}^{k} n_i \boldsymbol{S}_i$$

El caso en (b) es análogo (se deja como Ejercicio).

[2] Es decir $\widehat{\boldsymbol{\Sigma}}_{ii} = \widehat{\boldsymbol{\Sigma}}$, para $i = 1, \dots, k$.