MAT-269: Sesión 6
Estimación para la normal multivariada

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Estadisticas suficientes para μ y Σ

Resultado 1

Suponga $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ vectores aleatorios independientes siguiendo una distribución $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces $(\overline{\boldsymbol{x}}, \boldsymbol{Q})$ es estadística suficiente para $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con

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

Demostración:

La función de densidad conjunta para $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ adopta la forma:

$$\begin{aligned} f(\boldsymbol{X}) &= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\Big\{ -\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \Big\} \\ &= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\big\{ -\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} (\boldsymbol{Q} + n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^\top) \big\}, \end{aligned}$$

pertenece a la FE $\{p + p(p+1)/2\}$-paramétrica. Es decir, $(\overline{\boldsymbol{x}}, \boldsymbol{Q})$ o bien $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ son estadísticas suficientes para $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Resultado 2

Suponga $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ vectores aleatorios independientes cada uno $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $n > p$. Entonces los estimadores máximo verosímiles de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son

$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}, \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Q}.$$

EX UMBRA IN SOLEM

Estimación máximo verosímil (ML) de μ y Σ

Demostración:

Ignorando términos que no dependen de $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, tenemos que la función de log-verosimilitud es:

$$\ell(\boldsymbol{\theta}) = -\tfrac{n}{2} \log |\boldsymbol{\Sigma}| - \tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}(\boldsymbol{\mu}),$$

donde

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

Note que

$$\mathsf{d}\, \boldsymbol{Q}(\boldsymbol{\mu}) = -\sum_{i=1}^{n} \big\{ (\mathsf{d}\, \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top + (\boldsymbol{x}_i - \boldsymbol{\mu})(\mathsf{d}\, \boldsymbol{\mu})^\top \big\}$$

De este modo,

$$\begin{aligned}
\mathsf{d}_\mu\, \ell(\boldsymbol{\theta}) &= -\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\, \mathsf{d}\, \boldsymbol{Q}(\boldsymbol{\mu}) \\
&= \tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \textstyle\sum_{i=1}^{n} \big\{ (\mathsf{d}\, \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top + (\boldsymbol{x}_i - \boldsymbol{\mu})(\mathsf{d}\, \boldsymbol{\mu})^\top \big\} \\
&= \tfrac{n}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \big\{ (\mathsf{d}\, \boldsymbol{\mu})(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^\top + (\overline{\boldsymbol{x}} - \boldsymbol{\mu})(\mathsf{d}\, \boldsymbol{\mu})^\top \big\} \\
&= \tfrac{n}{2} \{ (\overline{\boldsymbol{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathsf{d}\, \boldsymbol{\mu}) + (\mathsf{d}\, \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \} \\
&= n (\mathsf{d}\, \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}),
\end{aligned}$$

y el diferencial es cero si, $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$.

Recuerde que (ver Magnus y Neudecker, 2007)

$$\mathsf{d}\,\boldsymbol{F}^{-1} = -\boldsymbol{F}^{-1}(\mathsf{d}\,\boldsymbol{F})\boldsymbol{F}^{-1}, \qquad \mathsf{d}\log|\boldsymbol{F}| = \operatorname{tr}\boldsymbol{F}^{-1}\,\mathsf{d}\,\boldsymbol{F}.$$

De este modo,

$$\begin{aligned}
\mathsf{d}_{\Sigma}\,\ell(\boldsymbol{\theta}) &= -\frac{n}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} - \frac{1}{2}\operatorname{tr}\mathsf{d}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}(\boldsymbol{\mu}) \\
&= -\frac{n}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} + \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}(\boldsymbol{\mu}) \\
&= \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}(\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} - n\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma}\right\} \\
&= \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{Q}(\boldsymbol{\mu}) - n\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma},
\end{aligned}$$

y por tanto el primer diferencial es cero, si:[1]

$$\boldsymbol{Q}(\widehat{\boldsymbol{\mu}}) - n\widehat{\boldsymbol{\Sigma}} = \boldsymbol{0}, \qquad \text{es decir} \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Q}.$$

[1] Note que $\boldsymbol{Q}(\widehat{\boldsymbol{\mu}}) = \boldsymbol{Q}(\overline{\boldsymbol{x}}) \equiv \boldsymbol{Q}$.

Para apreciar que $\widehat{\boldsymbol{\mu}}$ y $\widehat{\boldsymbol{\Sigma}}$ son máximos, note que:

$$\begin{aligned}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) &= \operatorname{tr}\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top \\ &= \operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}+n\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})^\top \\ &= \operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}+n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}-\boldsymbol{\mu}).\end{aligned}$$

De este modo, la parte relevante de la log-verosimilitud asume la forma:

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

como $\boldsymbol{\Sigma}>0$ (y de ahí que $\boldsymbol{\Sigma}^{-1}>0$), sigue que:

$$n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}-\boldsymbol{\mu})\geq 0,$$

con la igualdad, si y solo si $\boldsymbol{\mu}=\overline{\boldsymbol{x}}$, es decir $\widehat{\boldsymbol{\mu}}=\overline{\boldsymbol{x}}$ corresponde al MLE para $\boldsymbol{\mu}$.

Por lo tanto,

$$\ell(\widehat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}) = \ell(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}) = -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q},$$

como $|\boldsymbol{\Sigma}^{-1}| = |\boldsymbol{\Sigma}^{-1} \boldsymbol{Q} \boldsymbol{Q}^{-1}| = |\boldsymbol{\Sigma}^{-1} \boldsymbol{Q}| |\boldsymbol{Q}|^{-1}$, obtenemos

$$\ell(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}) = \frac{n}{2} \log |\boldsymbol{\Sigma}^{-1} \boldsymbol{Q}| - \frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q} - \frac{n}{2} \log |\boldsymbol{Q}|.$$

Además,

$$|\boldsymbol{\Sigma}^{-1} \boldsymbol{Q}| = |\boldsymbol{\Sigma}^{-1} \boldsymbol{Q}^{1/2} \boldsymbol{Q}^{1/2}| = |\boldsymbol{Q}^{1/2}| |\boldsymbol{\Sigma}^{-1}| |\boldsymbol{Q}^{1/2}| = |\boldsymbol{Q}^{1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}^{1/2}|,$$

y

$$\operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q} = \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}^{1/2} \boldsymbol{Q}^{1/2} = \operatorname{tr} \boldsymbol{Q}^{1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}^{1/2},$$

es decir

$$\ell(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}) = \frac{n}{2} \log |\boldsymbol{Q}^{1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}^{1/2}| - \frac{1}{2} \operatorname{tr} \boldsymbol{Q}^{1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}^{1/2} - \frac{n}{2} \log |\boldsymbol{Q}|.$$

Sea $\lambda_1, \dots, \lambda_p$ los valores propios de $\boldsymbol{Q}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}^{1/2}$ (o bien de $\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}$), entonces

$$\begin{aligned}\ell(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}) &= \frac{n}{2} \log \prod_{i=1}^{p} \lambda_i - \frac{1}{2} \sum_{i=1}^{p} \lambda_i - \frac{n}{2} \log |\boldsymbol{Q}| \\ &= \frac{1}{2} \sum_{i=1}^{p} \big(n \log \lambda_i - \lambda_i\big) - \frac{n}{2} \log |\boldsymbol{Q}|.\end{aligned}$$

Dado que la función

$$g(\lambda) = n \log \lambda - \lambda,$$

tiene un único máximo en $\lambda = n$, sigue que

$$\ell(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}) \leq \frac{p}{2}(n \log n - n) - \frac{n}{2} \log |\boldsymbol{Q}|,$$

con la igualdad si $\lambda_i = n$ $(i = 1, \dots, p)$. Esta última condición es equivalente a

$$\boldsymbol{Q}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}^{1/2} = n\boldsymbol{I}_p.$$

De ahí que

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Q},$$

Finalmente se concluye que

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq n^{np/2} e^{-np/2} |\boldsymbol{Q}|^{-1/2},$$

con la igualdad si $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$ y $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Q}$, lo que finaliza la prueba.

Resultado 3

Sea $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ vectores aleatorios IID, tales que $\boldsymbol{x}_i \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, para $i = 1, \dots, n$, donde $\boldsymbol{\Sigma} > 0$. La matriz de informacion de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, (\operatorname{vech} \boldsymbol{\Sigma})^\top)^\top$ es dada por:

$$\boldsymbol{\mathcal{F}}_n(\boldsymbol{\theta}) = n \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{2} \boldsymbol{D}_p^\top (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \boldsymbol{D}_p \end{pmatrix},$$

donde $\operatorname{vech}(\cdot)$ denota la vectorización de todos los elementos diferentes de $\boldsymbol{\Sigma}$ y $\boldsymbol{D}_p$ es la matriz de duplicación de orden p.

Definición 1 (matriz de duplicación)

Para $\boldsymbol{A}$ matriz simétrica $p \times p$, sea $\operatorname{vech}(\boldsymbol{A})$ la vectorización de los elementos distintos de $\boldsymbol{A}$[2]. Existe una única matriz $\boldsymbol{D}_p \in \mathbb{R}^{p^2 \times p(p+1)/2}$ que transforma $\operatorname{vech}(\boldsymbol{A})$ en $\operatorname{vec}(\boldsymbol{A})$, es decir:

$$\boldsymbol{D}_p \operatorname{vech}(\boldsymbol{A}) = \operatorname{vec}(\boldsymbol{A}), \qquad (\boldsymbol{A} = \boldsymbol{A}^\top),$$

y análogamente,

$$\operatorname{vech}(\boldsymbol{A}) = \boldsymbol{D}_p^+ \operatorname{vec}(\boldsymbol{A}), \qquad \boldsymbol{D}_p^+ = (\boldsymbol{D}_p^\top \boldsymbol{D}_p)^{-1} \boldsymbol{D}_p^\top.$$

[2] En efecto, tenemos $p(p+1)/2$ elementos distintos.

Matriz de duplicación D_p

Ejemplo:

Considere una matriz 3×3,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

De este modo,

$$\operatorname{vech} \boldsymbol{A} = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}, \quad \boldsymbol{D}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

La función `duplication` disponible en la biblioteca MVT permite obtener $\boldsymbol{D}_p$.[3]

[3] URL: `http://mvt.mat.utfsm.cl`.

Demostración del Resultado 3:

Diferenciando $\mathsf{d}_\mu\, \ell(\boldsymbol{\theta})$ con relación a $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, obtenemos

$$\begin{aligned}
\mathsf{d}^2_\mu\, \ell(\boldsymbol{\theta}) &= n(\mathsf{d}\,\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}\,\mathsf{d}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \\
&= -n(\mathsf{d}\,\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\mu}, \\
\mathsf{d}^2_{\Sigma\mu}\, \ell(\boldsymbol{\theta}) &= n(\mathsf{d}\,\boldsymbol{\mu})^\top (\mathsf{d}\,\boldsymbol{\Sigma}^{-1})(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \\
&= -n(\mathsf{d}\,\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}).
\end{aligned}$$

Ahora, diferenciando $\mathsf{d}_\Sigma\, \ell(\boldsymbol{\theta})$ con relación a $\boldsymbol{\Sigma}$ se tiene que

$$\begin{aligned}
\mathsf{d}^2_\Sigma\, \ell(\boldsymbol{\theta}) &= -\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\{\boldsymbol{Q}(\boldsymbol{\mu}) - n\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} \\
&\quad + \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\{\boldsymbol{Q}(\boldsymbol{\mu}) - n\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} \\
&\quad - \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{Q}(\boldsymbol{\mu}) - n\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} \\
&= -\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{Q}(\boldsymbol{\mu}) - n\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} \\
&\quad - \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{Q}(\boldsymbol{\mu}) - n\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} \\
&\quad - \tfrac{n}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma}.
\end{aligned}$$

Matriz de información de Fisher

Recordando que $\mathsf{E}(\overline{\boldsymbol{x}}) = \boldsymbol{\mu}$ y $\mathsf{E}\{\boldsymbol{Q}(\boldsymbol{\mu})\} = n\boldsymbol{\Sigma}$, sigue:

$$\begin{aligned}
\mathsf{E}\{-\,\mathsf{d}^2_{\mu}\,\ell(\boldsymbol{\theta})\} &= n(\mathsf{d}\,\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\mu}, \\
\mathsf{E}\{-\,\mathsf{d}^2_{\Sigma\mu}\,\ell(\boldsymbol{\theta})\} &= \boldsymbol{0} \\
\mathsf{E}\{-\,\mathsf{d}^2_{\Sigma}\,\ell(\boldsymbol{\theta})\} &= \frac{n}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma}.
\end{aligned}$$

Note que podemos escribir:

$$\begin{aligned}
\frac{n}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} &= \frac{n}{2}(\operatorname{vec}\boldsymbol{\Sigma})^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\operatorname{vec}\boldsymbol{\Sigma} \\
&= \frac{n}{2}(\operatorname{vech}\boldsymbol{\Sigma})^\top \boldsymbol{D}_p^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_p\operatorname{vech}\boldsymbol{\Sigma}.
\end{aligned}$$

De ahí sigue que:[4]

$$\mathsf{E}\left\{-\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\,\partial\boldsymbol{\mu}^\top}\right\} = n\boldsymbol{\Sigma}^{-1}, \qquad \mathsf{E}\left\{-\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\,\partial(\operatorname{vech}\boldsymbol{\Sigma})^\top}\right\} = \boldsymbol{0},$$

$$\mathsf{E}\left\{-\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\operatorname{vech}\boldsymbol{\Sigma}\,\partial(\operatorname{vech}\boldsymbol{\Sigma})^\top}\right\} = \frac{n}{2}\boldsymbol{D}_p^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_p.$$

[4] Usando el 2do Teorema de identificación de Magnus y Neudecker (2007).

Resultado 4

Sea $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ una muestra aleatoria desde una distribución $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces $\overline{\boldsymbol{x}}$ y $\widehat{\boldsymbol{\Sigma}}$ son estimadores consistentes[5] de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, respectivamente. Además, $\overline{\boldsymbol{x}}$ y $\widehat{\boldsymbol{\Sigma}}$ son asintóticamente independientes con distribuciones

$$\sqrt{n}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

$$\sqrt{n}(\operatorname{vech} \widehat{\boldsymbol{\Sigma}} - \operatorname{vech} \boldsymbol{\Sigma}) \xrightarrow{\mathsf{D}} \mathsf{N}_{p^*}(\boldsymbol{0}, 2\boldsymbol{D}_p^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})(\boldsymbol{D}_p^+)^\top),$$

con $p^* = p(p+1)/2$.

[5] Es decir, $\overline{\boldsymbol{x}} \xrightarrow{\mathsf{P}} \boldsymbol{\mu}$ y $\widehat{\boldsymbol{\Sigma}} \xrightarrow{\mathsf{P}} \boldsymbol{\Sigma}$.

Casos particulares (Tarea)

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

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:

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

Casos particulares (Tarea)

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,

$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}, \qquad \widehat{\phi} = \frac{1}{p}\operatorname{tr}(\boldsymbol{V}^{-1}\boldsymbol{S}).$$