ESTADÍSTICA MULTIVARIANTE

UGR, GRADO EN MATEMÁTICAS
Curso Académico 2023-2024

José Miguel Angulo Ibáñez (*jmangulo@ugr.es*)

Departamento de Estadística e Investigación Operativa
Universidad de Granada

- **TEMA 2. Inferencia en la Distribución Normal Multivariante**

  Estimadores de máxima verosimilitud de $\boldsymbol{\mu}$ y Σ en una DNM (II): Propiedades. Teorema de Zehna. Complemento: Distribución asintótica de $\bar{\mathbf{X}}$ y A para poblaciones no necesariamente normales

En este apartado se trata, a continuación, sobre el posible cumplimiento de algunas propiedades importantes por los EMV $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ y $\hat{\Sigma} = S_N$. Concretamente:

- INSESGADEZ
- CONSISTENCIA (fuerte, débil)
- EFICIENCIA

DEFINICIÓN: Sea $\{\mathbf{X}_\alpha : \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de una distribución dependiente de un parámetro (en general, multidimensional), $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \in \Theta \subseteq \mathbb{R}^k$. Un estimador $\hat{\boldsymbol{\theta}}$ (*i. e.*, función medible de la muestra, $\hat{\boldsymbol{\theta}} = \mathbf{T}(\mathrm{X})$) del parámetro $\boldsymbol{\theta}$ se dice *insesgado* si

$$E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \mathbf{0}, \quad \forall \boldsymbol{\theta} \in \Theta \qquad \text{(es decir, } E_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}, \quad \forall \boldsymbol{\theta} \in \Theta\text{)}$$

Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$

- VECTOR DE MEDIAS MUESTRAL:

$\boxed{\hat{\boldsymbol{\mu}} := \bar{\mathbf{X}} \text{ es un estimador insesgado de } \boldsymbol{\mu}}$ (□ Probar)

(En realidad, en este caso no hace falta suponer normalidad, solo la existencia de $\boldsymbol{\mu}$)

- MATRIZ DE COVARIANZAS MUESTRAL:

$\boxed{\hat{\Sigma} := S_N \quad \textbf{no} \text{ es un estimador insesgado de } \Sigma}$ (□ Probar)

(Tener en cuenta, por el teorema de Fisher, que $A \stackrel{d}{=} \sum_{\alpha=1}^{N-1} \mathbf{Z}_\alpha \mathbf{Z}'_\alpha$, con $\mathbf{Z}_\alpha \sim N_p(\mathbf{0}, \Sigma)$, $\alpha = 1, \ldots, N-1$, independientes)

Se tiene, en este caso, que

$$\boxed{S_{N-1} = \frac{N}{N-1} S_N \quad \textbf{sí} \text{ es un estimador insesgado de } \Sigma}$$

DEFINICIÓN: Sea $\{\mathbf{X}_\alpha : \ \alpha = 1, \ldots, N\}$ una muestra aleatoria simple (de tamaño N, considerado variable) de una distribución dependiente de un parámetro (en general, multidimensional), $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \in \Theta \subseteq \mathbb{R}^k$. Un estimador $\hat{\boldsymbol{\theta}}_N$ (*i. e.*, función medible de la muestra, $\hat{\boldsymbol{\theta}}_N = \mathbf{T}_N(\mathrm{X})$) del parámetro $\boldsymbol{\theta}$ se dice

(a) *débilmente consistente* si $\hat{\boldsymbol{\theta}}_N$ converge en probabilidad a $\boldsymbol{\theta}$, es decir,

$$\forall \epsilon > 0, \quad \lim_{N\to\infty} P_{\boldsymbol{\theta}} \left[\|\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\| < \epsilon \right] = 1$$

(b) *fuertemente consistente* si $\hat{\boldsymbol{\theta}}_N$ converge casi seguramente a $\boldsymbol{\theta}$, es decir,

$$P_{\boldsymbol{\theta}} \left[\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta} \right] = 1$$

Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$. Se demuestra que

$\hat{\boldsymbol{\mu}} := \bar{\mathbf{X}}$ y $\hat{\Sigma} := S_N$ son estimadores **fuertemente** consistentes de $\boldsymbol{\mu}$ y Σ, respectivamente

(En el caso de $\hat{\boldsymbol{\mu}} := \bar{\mathbf{X}}$, no se requiere normalidad. S_{N-1} también es un estimador fuertemente consistente de Σ) [Se omite la demostración]

(Para estimadores insesgados)

DEFINICIÓN: Sea $\mathbf{T} = (T_1, \ldots, T_k)'$ un estimador insesgado de un parámetro (en general, multidimensional) $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \in \Theta$, siendo $\Theta = \mathbb{R}^k$ o un 'rectángulo' en $\mathbb{R}^k$. Se dice que $\mathbf{T}$ es *eficiente* para $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ si, para cualquier otro estimador insesgado $\mathbf{U}$ de $\boldsymbol{\theta}$, se verifica que la diferencia de matrices de covarianzas $\mathrm{Cov}_{\boldsymbol{\theta}}(\mathbf{U}) - \mathrm{Cov}_{\boldsymbol{\theta}}(\mathbf{T})$, para todo $\boldsymbol{\theta} \in \Theta$, es una matriz definida no negativa [se suele denotar $\mathrm{Cov}_{\boldsymbol{\theta}}(\mathbf{U}) \geq \mathrm{Cov}_{\boldsymbol{\theta}}(\mathbf{T})$]

Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$. Se demuestra que:

- $\boxed{\bar{\mathbf{X}} \text{ es un estimador eficiente de } \boldsymbol{\mu} \text{ en el espacio } \mathbb{R}^p}$
- $\boxed{S_{N-1}$ es un estimador eficiente de Σ en el espacio de matrices simétricas definidas positivas de dimensión $p \times p}$

TEOREMA (Zehna, 1966): Sea $\mathcal{P} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ una familia de distribuciones de probabilidad sobre el espacio $(\mathbb{R}^p, \mathcal{B}^p)$. Sea $g : \Theta \rightarrow \Lambda$ una función arbitraria dada. Si $\hat{\boldsymbol{\theta}}$ es un estimador máximo-verosímil de $\boldsymbol{\theta}$, entonces $g(\hat{\boldsymbol{\theta}})$ es un estimador máximo-verosímil de $g(\boldsymbol{\theta})$

[OBSERVACIÓN: Este teorema hace uso del concepto de 'función de verosimilitud *inducida*' por g: Para cada $\lambda \in g(\Theta) \subseteq \Lambda$, se define

$$M(\lambda) = \sup_{\boldsymbol{\theta} \in \Theta_\lambda} L(\boldsymbol{\theta}), \qquad \text{con} \quad \Theta_\lambda = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) = \lambda\}$$

Se entiende entonces que '$g(\hat{\boldsymbol{\theta}})$ es un EMV de $g(\boldsymbol{\theta})$' en el sentido de que $g(\hat{\boldsymbol{\theta}})$ es un máximo para M]

En el contexto de la inferencia sobre la DNM, se tiene que para distintos coeficientes que dependen funcionalmente de $\boldsymbol{\mu}$ y Σ (por ejemplo, coeficientes de correlación, regresión, etc.) los correspondientes estimadores máximo-verosímiles se obtienen mediante sustitución de $\boldsymbol{\mu}$ y Σ por $\bar{\mathbf{X}}$ y S_N, respectivamente, en las expresiones que definen dichos coeficientes

[Complemento] Distribuciones asintóticas de $\hat{\mu}$ y $\hat{\Sigma}$, con distribución de referencia no necesariamente normal

- MOTIVACIÓN (referida al vector de medias muestral):

  Recordemos que, al enunciar y probar el teorema de Fisher, se ha obtenido que para una muestra aleatoria simple $\mathbf{X}_1, \ldots, \mathbf{X}_N$ de una distribución $N_p(\boldsymbol{\mu}, \Sigma)$ se tiene que

  $$\bar{\mathbf{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{\Sigma}{N}\right);$$

  es decir, normalizando en media y tamaño muestral,

  $$N^{\frac{1}{2}}(\bar{\mathbf{X}} - \boldsymbol{\mu}) = N^{-\frac{1}{2}} \sum_{\alpha=1}^{N} (\mathbf{X}_\alpha - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \Sigma)$$

  (la distribución límite ya no depende de N).

- ¿Qué puede afirmarse, al menos como aproximación límite (para $N \to \infty$), cuando la distribución de origen no es necesariamente una DNM?

- **Distribución asintótica del vector de medias muestral, $\bar{\mathbf{X}}$**

RESULTADO: Sea $\mathbf{X}_1, \ldots, \mathbf{X}_N, \ldots$ una sucesión de vectores aleatorios p-dimensionales independientes e idénticamente distribuidos, con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas Σ (*i. e.*, $\mathbf{X}_\alpha \sim (\boldsymbol{\mu}, \Sigma)$, $\alpha = 1, \ldots, N, \ldots$). Sea $\bar{\mathbf{X}}_N = \frac{1}{N} \sum_{\alpha=1}^{N} \mathbf{X}_\alpha$, $\forall N \geq 1$. Entonces, cuando $N \to \infty$, se tiene la distribución asintótica

$$N^{\frac{1}{2}}(\bar{\mathbf{X}}_N - \boldsymbol{\mu}) = N^{-\frac{1}{2}} \sum_{\alpha=1}^{N} (\mathbf{X}_\alpha - \boldsymbol{\mu}) \underset{N\to\infty}{\sim} N_p(\mathbf{0}, \Sigma)$$

(Se trata de una versión del Teorema Central del Límite, para vectores i.i.d. $\sim (\boldsymbol{\mu}, \Sigma)$)

- **Distribución asintótica de la matriz de dispersiones muestral, A**

RESULTADO: Sea $\mathbf{X}_1, \ldots, \mathbf{X}_N, \ldots$ una sucesión de vectores aleatorios p-dimensionales independientes e idénticamente distribuidos, con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas Σ (*i. e.*, $\mathbf{X}_\alpha \sim (\boldsymbol{\mu}, \Sigma)$, $\alpha = 1, \ldots, N, \ldots$), y con momentos de orden cuatro finitos. Sean $\bar{\mathbf{X}}_N = \frac{1}{N}\sum_{\alpha=1}^N \mathbf{X}_\alpha$ y $A_N = \sum_{\alpha=1}^N (\mathbf{X}_\alpha - \bar{\mathbf{X}}_N)(\mathbf{X}_\alpha - \bar{\mathbf{X}}_N)'$, $\forall N \geq 1$. Entonces, cuando $N \to \infty$, se tiene la distribución asintótica

$$N^{-\frac{1}{2}}(A_N - N\Sigma) \underset{N\to\infty}{\sim} N_{p^2}(\mathbf{0}, V)$$

(en el sentido de que $N^{-\frac{1}{2}}(\text{Vec}(A_N) - N\text{Vec}(\Sigma)) \underset{N\to\infty}{\sim} N_{p^2}(\mathbf{0}, V)$), con $V = \text{Cov}(\text{Vec}((\mathbf{X}_\alpha - \boldsymbol{\mu})(\mathbf{X}_\alpha - \boldsymbol{\mu})'))$

(OBSERVACIÓN: La matriz V será necesariamente singular, por lo que la distribución límite se refiere al caso general de la DNM)