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 (I): Obtención. Teorema de Dykstra. Teorema de Fisher multivariante

Función de verosimilitud

- Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$, y sea $\{\mathbf{X}_\alpha : \ \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de dicha distribución.

  Para una realización dada de la muestra, $\{\mathbf{x}_\alpha : \ \alpha = 1, \ldots, N\}$, la *función de verosimilitud* se expresa como la función (de argumentos $\boldsymbol{\mu}$ y Σ)

$$L(\boldsymbol{\mu}, \Sigma; \mathbf{x}_1, \ldots, \mathbf{x}_N) := f_{\boldsymbol{\mu},\Sigma}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \prod_{\alpha=1}^{N} f_{\boldsymbol{\mu},\Sigma}(\mathbf{x}_\alpha)$$

$$= \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}_\alpha - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_\alpha - \boldsymbol{\mu})\right\}$$

$$= \frac{1}{(2\pi)^{\frac{pN}{2}} |\Sigma|^{\frac{N}{2}}} \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_\alpha - \boldsymbol{\mu})\right\}$$

  De forma abreviada, se denotará simplemente $L(\boldsymbol{\mu}, \Sigma)$, sobreentendiéndose implícita la realización muestral de referencia

Función de verosimilitud (cont.)

- Mediante operaciones algebraicas, y usando la fórmula de momentos multivariante generalizada (caso $\mathbf{b} = \boldsymbol{\mu}$), se comprueba que la función de verosimilitud puede expresarse de la forma

$$L(\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{pN}{2}} |\Sigma|^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2} \mathrm{tr}(\Sigma^{-1}\tilde{A}) - \frac{N}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})'\Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \right\}$$

con $\tilde{A} = \sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})'$ (es decir, la matriz de dispersiones muestral evaluada en la realización dada de la muestra, $\{\mathbf{x}_\alpha : \alpha = 1, \ldots, N\}$) (□ Probar)

Más convenientemente, a efectos de maximización, se suele usar la transformación

$$\ln(L(\boldsymbol{\mu}, \Sigma)) = -\frac{pN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}\tilde{A}) - \frac{N}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})'\Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$$

(Es habitual proceder, equivalentemente, a la minimización de la función $-\ln(L(\boldsymbol{\mu}, \Sigma))$)

- **RESULTADO:**

  Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$, y sea $\{\mathbf{X}_\alpha : \ \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de dicha distribución. Entonces, los *estimadores máximo-verosímiles* (EMV) de $\boldsymbol{\mu}$ y Σ son, respectivamente,

  $$\bar{\mathbf{X}} \qquad \text{y} \qquad \frac{A}{N} = S_N,$$

  este último bajo la condición de ser A definida positiva $(*)$

  [$(*)$ Este aspecto se discutirá más adelante, tras la demostración, en relación con el ‘Teorema de Dykstra’]

Estimadores máximo-verosímiles de μ y Σ (cont.)

- El problema de optimización se podría abordar de forma directa mediante derivación matricial. No obstante, es interesante probar el resultado por el procedimiento que se expodrá a continuación
- Para la demostración, se utilizará el siguiente resultado auxiliar en la parte relativa al estimador del parámetro Σ:

  LEMA (Watson): Sea

  $$f(G) = -N \ln(|G|) - \mathrm{tr}(G^{-1}D),$$

  con argumento G una matriz simétrica definida positiva, y siendo D una matriz simétrica definida positiva dada, ambas $p \times p$. Entonces, existe (y es único) el máximo de f respecto a G, y se alcanza en $G = \frac{1}{N}D$, siendo el valor máximo alcanzado

  $$f\left(\frac{1}{N}D\right) = pN \ln(N) - N \ln(|D|) - pN$$

  ($\square$ Probar)

- **Demostración**:

  - Maximización de ln$(L(\boldsymbol{\mu},\Sigma))$ en $\boldsymbol{\mu}$:

    (Se puede hacer, en este caso, independientemente de Σ)

    De la forma obtenida para ln$(L(\boldsymbol{\mu},\Sigma))$, se desprende que, independientemente del valor de Σ, el máximo se alcanzará donde se minimice la forma cuadrática

    $$(\bar{\mathbf{x}}-\boldsymbol{\mu})'\Sigma^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})$$

    Ahora bien, dado que Σ^{-1} es definida positiva (por serlo Σ), la forma cuadrática alcanza el valor mínimo 0 para (y solo para) $\bar{\mathbf{x}}-\boldsymbol{\mu}=0$, es decir, $\boldsymbol{\mu}=\bar{\mathbf{x}}$. Por tanto, el estadístico definido por

    $$\hat{\mu} := \bar{\mathbf{X}} \qquad \text{(el vector de medias muestral)}$$

    es el EMV (único, c.s.) de $\boldsymbol{\mu}$

- Maximización de $\ln(L(\bar{\mathbf{x}}, \Sigma))$ en Σ:

  Dado que tratamos de maximizar la función

  $$\ln(L(\bar{\mathbf{x}}, \Sigma)) = -\frac{pN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}\tilde{A})$$

  (donde ya ha desaparecido el término correspondiente a la forma cuadrática), identificamos, en la notación del 'lema de Watson',

  $$G := \Sigma, \qquad D := \tilde{A},$$

  $$f(\Sigma) := 2\left[\ln L(\bar{\mathbf{x}}, \Sigma) + \frac{pN}{2}\ln(2\pi)\right] = pN\ln(2\pi) + 2\ln(L(\bar{\mathbf{x}}, \Sigma))$$
  $$= -N\ln(|\Sigma|) - \mathrm{tr}(\Sigma^{-1}\tilde{A}),$$

  por lo que el máximo de f en Σ (igualmente, entonces, el máximo de $L(\bar{\mathbf{x}}, \Sigma)$ en Σ) se alcanza para (y solo para) $\Sigma = \frac{\tilde{A}}{N} =: \tilde{S_N}$. Por tanto, el estadístico definido por

  $$\hat{\Sigma} := \frac{A}{N} = S_N \qquad \text{(la matriz de covarianzas muestral)}$$

  es el EMV (único, c.s.) de Σ ■

Estimadores máximo-verosímiles de μ y Σ (cont.)

- Se comprueba que el valor máximo alcanzado por la función de verosimilitud en el punto $(\boldsymbol{\mu}, \Sigma) = (\bar{\mathbf{x}}, \tilde{S_N})$ del espacio paramétrico es

$$L(\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi e)^{\frac{pN}{2}} |\tilde{S_N}|^{\frac{N}{2}}} = \left[\frac{N}{(2\pi e)|\tilde{A}|^{\frac{1}{p}}} \right]^{\frac{pN}{2}}$$

(□ Probar)

Como puede verse, este valor máximo no depende directamente del vector de medias muestral, sino del determinante (producto de autovalores) de la matriz dispersiones muestral respecto a éste, de forma inversa

Teorema de Dykstra

- En relación con el resultado anterior sobre el EMV del par paramétrico $(\boldsymbol{\mu}, \Sigma)$, se plantea la cuestión siguiente:

  Formalmente, puede ocurrir que para una realización concreta de la muestra aleatoria la matriz $\tilde{A}$ resulte ser singular (es decir, semidefinida positiva). Ahora bien, ¿con qué probabilidad puede darse este caso?

- La respuesta viene dada por el 'teorema de Dykstra' (1970):

  TEOREMA (Dykstra):

  Sea $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$. Sea $\{\mathbf{X}_\alpha : \ \alpha = 1, \ldots, N\}$ una muestra aleatoria simple de dicha distribución y sea $A = \sum_{\alpha=1}^{N} (\mathbf{X}_\alpha - \bar{\mathbf{X}})(\mathbf{X}_\alpha - \bar{\mathbf{X}})'$ la correspondiente matriz de dispersiones muestral. Entonces, A es definida positiva, con probabilidad 1, si y solo si $N > p$.

- Para la demostración del teorema, se establecen algunos resultados previos:

LEMA I:

Sean $\{\mathbf{X}_\alpha : \alpha = 1, \ldots, N\}$ vectores [aleatorios o no] de dimensión p. Sea $C = \left(c_{\alpha\beta}\right)_{\alpha,\beta=1,\ldots,N}$ una matriz ortogonal de dimensión $N \times N$. Entonces, definiendo

$$\mathbf{Y}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}\mathbf{X}_\beta, \quad \alpha = 1, \ldots, N,$$

se tiene que

$$\sum_{\alpha=1}^{N} \mathbf{X}_\alpha \mathbf{X}'_\alpha = \sum_{\alpha=1}^{N} \mathbf{Y}_\alpha \mathbf{Y}'_\alpha$$

(□ Probar)

LEMA II:

Sean $\{\mathbf{X}_\alpha : \alpha = 1, \ldots, N\}$ vectores aleatorios de dimensión p, con $\mathbf{X}_\alpha \sim N_p(\boldsymbol{\mu}_\alpha, \Sigma)$, $\alpha = 1, \ldots, N$ (es decir, tienen la misma matriz de covarianzas, aunque posiblemente distinto vector de medias), independientes. Sea $C = \left(c_{\alpha\beta}\right)_{\alpha,\beta=1,\ldots,N}$ una matriz ortogonal de dimensión $N \times N$. Entonces, definiendo

$$\mathbf{Y}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}\mathbf{X}_\beta, \quad \alpha = 1, \ldots, N,$$

se tiene que

- $\mathbf{Y}_\alpha \sim N_p(\boldsymbol{\nu}_\alpha, \Sigma)$, con $\boldsymbol{\nu}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}\boldsymbol{\mu}_\beta, \quad \alpha = 1, \ldots, N$
- Los vectores $\{\mathbf{Y}_\alpha : \alpha = 1, \ldots, N\}$ son independientes

(□ Probar)

- **Demostración** (esquema):

  Consideremos, en nuestro caso, una matriz $B = \left(b_{\alpha\beta}\right)_{\alpha,\beta=1,\ldots,N}$ de dimensión $N \times N$, ortogonal, cuya última fila sea $(\frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}})$. Es decir, el resto de filas han de ser vectores ortogonales con éste y entre sí, con norma igual a 1. Definimos el vector

  $$\mathbf{Z}_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta}\mathbf{X}_\beta, \quad \alpha = 1, \ldots, N$$

  Por los resultados anteriores (lemas I y II), tenemos que

  - $\mathbf{Z}_\alpha \sim N_p(\boldsymbol{\nu}_\alpha, \Sigma)$, con

  $$\boldsymbol{\nu}_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta}\boldsymbol{\mu} = \left(\sum_{\beta=1}^{N} b_{\alpha\beta}\right)\boldsymbol{\mu} = \begin{cases} \mathbf{0}, & \text{si } \alpha = 1, \ldots, N-1, \\ \frac{N}{\sqrt{N}}\boldsymbol{\mu} = \sqrt{N}\boldsymbol{\mu}, & \text{si } \alpha = N \end{cases}$$

Teorema de Dykstra (cont.)

- Los vectores $\{\mathbf{Z}_\alpha : \ \alpha = 1, \ldots, N\}$ son <u>independientes</u>

- $\mathbf{Z}_N = \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}} \mathbf{X}_\beta = \sqrt{N} \left(\frac{1}{N} \sum_{\beta=1}^{N} \mathbf{X}_\beta \right) = \sqrt{N} \bar{\mathbf{X}}$

  (es decir, $\bar{\mathbf{X}} = \frac{1}{\sqrt{N}} \mathbf{Z}_N$)

- $A = \sum_{\alpha=1}^{N} (\mathbf{X}_\alpha - \bar{\mathbf{X}})(\mathbf{X}_\alpha - \bar{\mathbf{X}})' = \sum_{\alpha=1}^{N} \mathbf{X}_\alpha \mathbf{X}'_\alpha - N \bar{\mathbf{X}} \bar{\mathbf{X}}'$

  $= \sum_{\alpha=1}^{N} \mathbf{Z}_\alpha \mathbf{Z}'_\alpha - \mathbf{Z}_N \mathbf{Z}'_N = \sum_{\alpha=1}^{N-1} \mathbf{Z}_\alpha \mathbf{Z}'_\alpha$

- Como consecuencia, $\bar{\mathbf{X}}$ y A son <u>independientes</u>

Finalmente, observemos que podemos escribir

$$A = \mathrm{Z}'\mathrm{Z}, \quad \text{con} \quad \mathrm{Z} = \begin{pmatrix} \mathbf{Z}_1' \\ \vdots \\ \mathbf{Z}_{N-1}' \end{pmatrix}, \text{ matriz } (N-1) \times p,$$

$$\mathrm{Z}' = \begin{pmatrix} \mathbf{Z}_1 & \cdots & \mathbf{Z}_{N-1} \end{pmatrix}, \text{ matriz } p \times (N-1)$$

Esto equivale a que la matriz A es definida no negativa.

Por otra parte, se cumple que

$$\text{rango}(\mathrm{Z}) = \text{rango}(\mathrm{Z}') = \text{rango}(\mathrm{Z}'\mathrm{Z}) = \text{rango}(A)$$

Por tanto, la demostración quedará concluida probando que

$$\text{rango}(\mathrm{Z}) = p, \text{ con probabilidad 1, si y solo si } N > p$$

(□ Probar) ■

De la demostración del teorema de Dykstra, se tiene también la prueba del siguiente resultado, conocido como 'teorema de Fisher multivariante':

TEOREMA (Fisher):

Dada una muestra aleatoria simple $\mathbf{X}_1, \ldots, \mathbf{X}_N$ de una distribución $N_p(\boldsymbol{\mu}, \Sigma)$, el vector de medias muestral $\bar{\mathbf{X}}$ se distribuye según una $N_p\left(\boldsymbol{\mu}, \frac{\Sigma}{N}\right)$, y la matriz de dispersiones muestral de igual modo que $\sum_{\alpha=1}^{N-1} \mathbf{Z}_\alpha \mathbf{Z}'_\alpha$, con $\mathbf{Z}_1, \ldots, \mathbf{Z}_{N-1}$ vectores aleatorios independientes e idénticamente distribuidos según una $N_p(\mathbf{0}, \Sigma)$, siendo el vector $\bar{\mathbf{X}}$ y la matriz A independientes