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 1. Distribución Normal Multivariante**

  Extensión al caso general, $\Sigma \geq 0$

- En el apartado anterior se ha introducido la DNM en el caso $\Sigma > 0$, dándose varias formulaciones equivalentes, una de las cuales se ha adoptado como definición y las otras como caracterizaciones:
  - **[D]** En términos de la densidad asociada (se requiere formalmente la condición $|\Sigma| \neq 0$ y, por tanto, $\exists \Sigma^{-1}$)
  - **[C-I]** Como transformación lineal (no singular) de un vector de componentes independientes con DN univariante estándar
  - **[C-II]** En términos de la forma de la función característica
  - **[C-III]** Bajo la condición de normalidad (no degenerada) de combinaciones lineales de las componentes
  - **[C-IV]** Mediante transformaciones lineales (no singulares) a partir de la densidad esférica estándar y radio aleatorio $\sqrt{\chi_p^2}$ (donde también se requiere explícitamente la existencia de Σ^{-1})
- Con el fin de dar una formulación de la DNM extendida al caso general en que $\Sigma \geq 0$ (pudiendo ser singular, es decir, semidefinida positiva), se observa que [C-I], [C-II] y [C-III] ofrecen una viabilidad más directa, especialmente en el caso de [C-II]

Introducción

- En el apartado anterior se ha introducido la DNM en el caso $\Sigma > 0$, dándose varias formulaciones equivalentes, una de las cuales se ha adoptado como definición y las otras como caracterizaciones:

  [D] En términos de la densidad asociada (se requiere formalmente la condición $|\Sigma| \neq 0$ y, por tanto, $\exists \Sigma^{-1}$)

  [C-I] Como transformación lineal (no singular) de un vector de componentes independientes con DN univariante estándar

  [C-II] En términos de la forma de la función característica

  [C-III] Bajo la condición de normalidad (no degenerada) de combinaciones lineales de las componentes

  [C-IV] Mediante transformaciones lineales (no singulares) a partir de la densidad esférica estándar y radio aleatorio $\sqrt{\chi_p^2}$ (donde también se requiere explícitamente la existencia de Σ^{-1})

- En este apartado, se trata de introducir dicha formulación extendida partiendo de la identificación de la DNM por su función característica (enfoque [C-II]) y discutiendo las implicaciones con respecto a las otras formas de interpretación

Sobre la descomposición espectral de la matriz de covarianzas, $\Sigma \geq 0$

- Sea Σ la matriz de covarianzas de algún vector aleatorio p-dimensional. Por ser $\Sigma \geq 0$ y simétrica, puede factorizarse de la forma

$$\Sigma = H\Lambda H', \quad \text{con:} \quad \begin{array}{l} H \text{ una matriz } p \times p \text{ ortogonal } (HH' = H'H = I_p) \\ \Lambda \text{ una matriz } p \times p \text{ diagonal, cuyos elementos} \\ \quad \text{diagonales distintos son los autovalores de } \Sigma \end{array}$$

Sea $r = \text{rango}(\Sigma)$, $r \leq p$. Supongamos, por conveniencia, que las matrices se eligen de modo que

$$\Lambda = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \text{ con } D \text{ una matriz } r \times r \text{ no singular, escribiendo}$$

$$H = \begin{pmatrix} H_1 & H_2 \end{pmatrix}, \text{ con } H_1 \text{ una matriz } p \times r \text{ y } H_2 \text{ una matriz } p \times (p - r)$$

Se tiene entonces que

$$\Sigma = \begin{pmatrix} H_1 & H_2 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} = H_1 D H_1'$$

Función característica de la DNM (caso $\Sigma \geq 0$)

Introducimos formalmente la DNM en el caso general ($\Sigma \geq 0$) mediante la identificación de su función característica:

- **DEFINICIÓN (extensión de [C-II])**:

  Un vector aleatorio p-dimensional $\mathbf{X}$ con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\Sigma \geq 0$ tiene una DNM, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, si y solo si su función característica tiene la forma

  $$\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{ i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t} \right\}$$

- Antes de tratar las extensiones de [C-I] y [C-III] a partir de esta definición, veremos un resultado general sobre transformaciones lineales:

- RESULTADO:

  Sean

$$\begin{aligned} \mathbf{X} &\sim N_p(\boldsymbol{\mu}_\mathbf{X}, \Sigma_\mathbf{X}) \quad (\Sigma_\mathbf{X} \geq 0) \\ \mathbf{Y} &= B\mathbf{X} + \mathbf{b} \quad \text{con} \\ & \qquad B \ \text{ matriz } q \times p \text{ (cte.)} \quad (\Rightarrow \text{rango}(B) \leq \min\{q, p\}) \\ & \qquad \mathbf{b} \ \text{ vector } q \times 1 \text{ (cte.)} \end{aligned}$$

  Entonces, se tiene que

$$\mathbf{Y} \sim N_q(\boldsymbol{\mu}_\mathbf{Y}, \Sigma_\mathbf{Y}) = N_q(B\boldsymbol{\mu}_\mathbf{X} + \mathbf{b}, B\Sigma_\mathbf{X}B'),$$

  (siendo rango$(B\Sigma_\mathbf{X}B') \leq$ min$\{$rango(B), rango$(\Sigma)\}$) (□ Probar)

- Intuitivamente, en el caso en que la matriz Σ sea singular, la variabilidad del vector aleatorio $\mathbf{X}$, con DNM (versión extendida), podrá ser explicada (con probabilidad 1) en términos de un número de variables aleatorias, con DN univariante e independientes, igual al rango de Σ. A continuación, vemos la argumentación formal para llegar al enunciado de la correspondiente caracterización, como extensión de [C-I].

- A partir de la descomposición espectral de Σ, vista al principio, de la forma
$$\Sigma = H_1 D H_1',$$
y teniendo en cuenta que
$$HH' = \begin{pmatrix} H_1 & H_2 \end{pmatrix} \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} = H_1H_1' + H_2H_2',$$
podemos escribir la función característica de $\mathbf{X}$ como
$$\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}} = e^{i\mathbf{t}'HH'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'H_1DH_1'\mathbf{t}} = e^{i\mathbf{t}'H_1H_1'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'H_1DH_1'\mathbf{t}} e^{i\mathbf{t}'H_2H_2'\boldsymbol{\mu}}$$

- Esta expresión sugiere considerar el cambio de variables lineal

$$\mathbf{Y} := H'\mathbf{X} = \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} \mathbf{X} = \begin{pmatrix} H_1'\mathbf{X} \\ H_2'\mathbf{X} \end{pmatrix} =: \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix},$$

de modo que, por ser H ortogonal, se tiene también que $\mathbf{X} = H\mathbf{Y}$.

Los momentos de primer y segundo orden de $\mathbf{Y}$ vienen dados por

$$\boldsymbol{\mu}_{\mathbf{Y}} = H'\boldsymbol{\mu} = \begin{pmatrix} H_1'\boldsymbol{\mu} \\ H_2'\boldsymbol{\mu} \end{pmatrix} =: \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{Y}_1} \\ \boldsymbol{\mu}_{\mathbf{Y}_2} \end{pmatrix}$$

$$\Sigma_{\mathbf{Y}} = H'\Sigma H = H'H\Lambda H'H = \Lambda = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} =: \begin{pmatrix} \Sigma_{\mathbf{Y}_1} & 0 \\ 0 & \Sigma_{\mathbf{Y}_2} \end{pmatrix}$$

Es decir,

$$\mathbf{Y}_1 \sim N_r(H_1'\boldsymbol{\mu}, D)$$

$$\mathbf{Y}_2 \sim N_{p-r}(H_2'\boldsymbol{\mu}, 0) \quad \text{(es decir, la distribución degenerada en } H_2'\boldsymbol{\mu}\text{: } P[\mathbf{Y}_2 = H_2'\boldsymbol{\mu}] = 1)$$

- Por tanto, tenemos que la función característica de $\mathbf{X}$ se puede interpretar como

$$\phi_{\mathbf{X}}(\mathbf{t}) = E\left[e^{i\mathbf{t}'\mathbf{X}}\right] = E\left[e^{i\mathbf{t}'H\mathbf{Y}}\right] = \phi_{\mathbf{Y}}(H'\mathbf{t})$$

Con el cambio de variables

$$\mathbf{v} := H'\mathbf{t} = \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} \mathbf{t} = \begin{pmatrix} H_1'\mathbf{t} \\ H_2'\mathbf{t} \end{pmatrix} =: \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix},$$

podemos escribir:

$$\begin{aligned} \phi_{\mathbf{Y}}(\mathbf{v}) = \phi_{\mathbf{X}}(\mathbf{t}) &= e^{i\mathbf{t}'H_1H_1'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'H_1DH_1'\mathbf{t}} e^{i\mathbf{t}'H_2H_2'\boldsymbol{\mu}} \\ &= e^{i\mathbf{v}_1'\boldsymbol{\mu}_{\mathbf{Y}_1} - \frac{1}{2}\mathbf{v}_1'\Sigma_{\mathbf{Y}_1}\mathbf{v}_1} e^{i\mathbf{v}_2'\boldsymbol{\mu}_{\mathbf{Y}_2}} \\ &= \phi_{\mathbf{Y}_1}(\mathbf{v}_1)\phi_{\mathbf{Y}_2}(\mathbf{v}_2) \end{aligned}$$

Es decir, $\mathbf{Y}_1$ e $\mathbf{Y}_2$ son independientes.

- Ahora, puesto que $\mathbf{Y}_1 \sim N_r(H_1'\boldsymbol{\mu}, D)$, con $D > 0$, se tiene por [C-I] que $\mathbf{Y}_1$ puede representarse en distribución como

$$\mathbf{Y}_1 \stackrel{d}{=} D^{\frac{1}{2}}\mathbf{Z} + H_1'\boldsymbol{\mu}, \quad \text{con } \mathbf{Z} \sim N_r(\mathbf{0}, I_r);$$

para $\mathbf{Y}_2$ se puede escribir

$$\mathbf{Y}_2 \stackrel{P\text{-c.s.}}{=} H_2'\boldsymbol{\mu} \equiv 0\mathbf{Z} + H_2'\boldsymbol{\mu}, \quad \text{siendo aquí } 0 \text{ una matriz } (p-r) \times r$$

Conjuntamente (teniendo en cuenta que $\stackrel{P\text{-c.s.}}{=} \Rightarrow \stackrel{d}{=}$), podemos escribir

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} D^{\frac{1}{2}}\mathbf{Z} + H_1'\boldsymbol{\mu} \\ 0\mathbf{Z} + H_2'\boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} D^{\frac{1}{2}} \\ 0 \end{pmatrix} \mathbf{Z} + H'\boldsymbol{\mu}$$

Se deshace el cambio de variables multiplicando por H,

$$\mathbf{X} = H\mathbf{Y} \stackrel{d}{=} H \begin{pmatrix} D^{\frac{1}{2}} \\ 0 \end{pmatrix} \mathbf{Z} + HH'\boldsymbol{\mu} = H_1 D^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$$

Finalmente, denotando $A = H_1 D^{\frac{1}{2}}$, se tiene que

$$\mathbf{X} \stackrel{d}{=} A\mathbf{Z} + \boldsymbol{\mu}, \quad \text{con: } A \text{ matriz } p \times r \text{ (cte.) de rango } r$$

$$\mathbf{Z} \sim N_r(\mathbf{0}, I_r)$$

- Recíprocamente, es inmediato que si se cumple que

$$\mathbf{X} \stackrel{d}{=} A\mathbf{Z} + \boldsymbol{\mu}, \quad \text{con: } A \text{ matriz } p \times r \text{ (cte.) de rango } r$$
$$\mathbf{Z} \sim N_r(\mathbf{0}, I_r),$$

teniendo en cuenta la correspondencia biunívoca entre distribuciones y funciones características (teorema de inversión de Paul Lévy), ha de ser

$$\phi_{\mathbf{X}} \equiv \phi_{A\mathbf{Z}+\boldsymbol{\mu}}$$

Es decir, para todo $\mathbf{t} \in \mathbb{R}^p$,

$$\begin{aligned} \phi_{\mathbf{X}}(\mathbf{t}) = \phi_{A\mathbf{Z}+\boldsymbol{\mu}}(\mathbf{t}) &= E\left[e^{i\mathbf{t}'(A\mathbf{Z}+\boldsymbol{\mu})}\right] \\ &= e^{i\mathbf{t}'\boldsymbol{\mu}} E\left[e^{i(\mathbf{t}'A)\mathbf{Z}}\right] = e^{i\mathbf{t}'\boldsymbol{\mu}} \phi_{\mathbf{Z}}(A'\mathbf{t}) \\ &= e^{i\mathbf{t}'\boldsymbol{\mu}} e^{-\frac{1}{2}(\mathbf{t}'A)(A'\mathbf{t})} = e^{i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'(AA')\mathbf{t}} \end{aligned}$$

Por tanto, según la definición general de la DNM dada a partir de la extensión de [C-II],

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma), \quad \text{con } \Sigma_{\mathbf{X}} = AA' \geq 0$$

Se puede enunciar, como conclusión, la caracterización siguiente:

- **CARACTERIZACIÓN I (caso general)**:

  Un vector aleatorio p-dimensional $\mathbf{X}$ tiene una DNM, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ ($\Sigma \geq 0$), si y solo si

$$\begin{aligned} \mathbf{X} \ &\stackrel{d}{=} \ A\mathbf{Z} + \boldsymbol{\mu} \quad \text{con} \\ &\qquad A \ \text{ matriz } p \times r \text{ (cte.), rango } (A) = r,, \ AA' = \Sigma \\ &\qquad \mathbf{Z} \sim N_r(\mathbf{0}, I_r) \end{aligned}$$

En relación con la caracterización [C-III], el resultado extendido al caso general puede enunciarse de la forma siguiente:

- **CARACTERIZACIÓN III (caso general)**:

  Un vector aleatorio p-dimensional $\mathbf{X} = (X_1, \ldots, X_p)'$ tiene una DNM (versión extendida) si y solo si toda combinación de la forma

  $$\boldsymbol{\alpha}'\mathbf{X}, \quad \text{con } \boldsymbol{\alpha} \in \mathbb{R}^p,$$

  tiene una DN univariante (posiblemente degenerada, *i. e.* con varianza nula)

  (□ Probar)

  (La demostración es similar a la correspondiente al caso no singular, con mínimas modificaciones)

Discusión en relación con la posible no existencia de densidad en todo $\mathbb{R}^p$ (caso $\Sigma \geq 0$)

- En referencia a la definición basada en la función de densidad, dada para el caso no singular ($\Sigma > 0$), en el caso general que nos ocupa ($\Sigma \geq 0$) sólo existirá propiamente la densidad para un subvector, con dimensión igual al rango de la matriz Σ, de un vector obtenido a partir de $\mathbf{X}$ mediante un cambio de variables conveniente

  Dicho cambio estará asociado a la descomposición espectral de Σ, y el subvector corresponderá a las componentes generadas por los autovectores asociados a los autovalores estrictamente positivos de Σ

  Es decir, se tiene que la distribución es no singular en un subespacio afín de dimensión $r = \text{rango}\,(\Sigma)$. Para el resto de dimensiones, la distribución es degenerada

- Análogamente, se podrá dar una representación como distribución simétrica de contornos elípticos restringida al subespacio afín de dimensión $r = \text{rango}\,(\Sigma)$ mencionado

- **RESULTADO**:

  Sean $\mathbf{X}_k$, $k = 1, \ldots, m$, vectores aleatorios p-dimensionales con DNM,

  $$\mathbf{X}_k \sim N_p(\boldsymbol{\mu}_k, \Sigma_k) \quad \text{(respectivamente)},$$

  e independientes. Entonces, para cualquier conjunto de matrices A_k, $k = 1, \ldots, m$ (ctes.), de dimensión $q \times p$, se verifica que

  $$\mathbf{Y} := \sum_{k=1}^{m} A_k \mathbf{X}_k \sim N_q\left(\sum_{k=1}^{m} A_k \boldsymbol{\mu}_k, \sum_{k=1}^{m} A_k \Sigma_k A_k'\right)$$

  (□ Probar)

[Complemento] El 'problema de Lévy-Cramér' (teorema de Cramér)

- Como consecuencia del resultado anterior, se tiene, en particular, el siguiente caso simple:

  Sean $\mathbf{X}_1$ y $\mathbf{X}_2$ vectores aleatorios p-dimensionales con

  $$\mathbf{X}_1 \sim N_p(\boldsymbol{\mu}_1, \Sigma_1) \qquad \mathbf{X}_2 \sim N_p(\boldsymbol{\mu}_2, \Sigma_2),$$

  e independientes. Entonces,

  $$\mathbf{X}_1 + \mathbf{X}_2 \sim N_p(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \Sigma_1 + \Sigma_2)$$

- En este caso (*i. e.* bajo normalidad), se tiene que el *recíproco* también es cierto:

  Sean $\mathbf{X}_1$ y $\mathbf{X}_2$ vectores aleatorios p-dimensionales independientes tales que $\mathbf{X}_1 + \mathbf{X}_2$ tiene DNM. Entonces, $\mathbf{X}_1$ y $\mathbf{X}_2$ tienen también, cada uno, DNM

[Complemento] El 'problema de Lévy-Cramér' (teorema de Cramér) (cont.)

- Históricamente, el resultado fue conjeturado por Paul Lévy, aunque no fue demostrado formalmente, por Harald Cramér, hasta 1936, mediante herramientas de la teoría de funciones analíticas, para el caso $p = 1$

- No veremos aquí la demostración. No obstante, es fácil probar (por ejemplo, mediante la caracterización [C-III]) que, partiendo de que el resultado es cierto para $p = 1$, lo es inmediatamente para cualquier p (□ Probar)