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**
  Caso $\Sigma > 0$: definiciones, propiedades y caracterizaciones

- DEFINICIÓN: Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio Se dice que $\mathbf{X}$ tiene una *distribución normal p-variante* si su densidad es de la forma

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

  siendo $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)' \in \mathbb{R}^p$ y Σ una matrix escalar ($p \times p$)-dimensional simétrica y definida positiva

  Se denota $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, representando $\boldsymbol{\mu}$ y Σ los 'parámetros' de la distribución

  Se demuestra que $f_{\mathbf{X}}$ está bien definida, es decir, que cumple las condiciones para ser una función de densidad:

  1. $f_{\mathbf{X}}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^p$ (□ Probar)
  2. $\int_{\mathbb{R}^p} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1$ (□ Probar)

- A partir de cualquier factorización de la forma

$$\Sigma = CC',$$

para alguna matriz C de dimensión $p \times p$ no singular (por ser $\Sigma > 0$), se tiene que el vec.a. $\mathbf{Z} = (Z_1, \ldots, Z_p)'$ definido por

$$\mathbf{Z} = C^{-1}(\mathbf{X} - \boldsymbol{\mu}) \qquad \text{(es decir, con } \mathbf{X} = C(\mathbf{Z} + \boldsymbol{\mu})\text{)}$$

se distribuye con función de densidad

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi}} \text{exp} \left\{ -\frac{1}{2} z^{\frac{1}{2}} \right\} = \prod_{j=1}^{p} f_{Z_j}(z_j)$$

Es decir, para cada $j = 1, \ldots, p$,

$$Z_j \sim N_1(0, 1),$$

siendo $Z_1, \ldots, Z_p$ independientes (por tanto, incorreladas), y

$$\mathbf{Z} \sim N_p(\mathbf{0}, I_p)$$

Vector de medias y matriz de covarianzas (cont.)

- Se comprueba de forma inmediata que $\boldsymbol{\mu}$ y Σ representan, respectivamente, el vector de medias y la matriz de covarianzas del vector aleatorio $\mathbf{X}$, es decir:

$$\boldsymbol{\mu}_{\mathbf{X}} = \boldsymbol{\mu}$$

$$\Sigma_{\mathbf{X}} = \Sigma$$

(□ Probar)

[1] CAMBIO DE VARIABLES LINEAL:

Sean

$$\begin{aligned} \mathbf{X} &\sim N_p(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}}) \quad (\Sigma > 0) \\ \mathbf{Y} &= B\mathbf{X} + \mathbf{b} \quad \text{con} \end{aligned}$$

B matriz $p \times p$ (cte.) no singular

$\mathbf{b}$ vector $p \times 1$ (cte.)

Entonces, se tiene que

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

(□ Probar)

- **CARACTERIZACIÓN I**:

Un vector aleatorio p-dimensional $\mathbf{X}$ tiene una DNM no singular, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ $(\Sigma > 0)$, si y solo si

$$\mathbf{X} \stackrel{d}{=} A\mathbf{Z} + \boldsymbol{\mu} \quad \text{con}$$

A matriz $p \times p$ (cte.) no singular, $AA' = \Sigma$

$\mathbf{Z} \sim N_p(\mathbf{0}, I_p)$

(□ Probar)

[2] INDEPENDENCIA Y CONDICIONAMIENTO:

Vemos a continuación algunos resultados fundamentales (dos sobre independencia y uno sobre condicionamiento) que, entre otras posibilidades, pueden tratarse a partir de la definición de la DNM no singular en términos de su densidad

- RESULTADO 1: Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio con DNM no singular,

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

Si la matriz Σ es diagonal,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix} = \text{diag}\,(\sigma_1^2, \ldots, \sigma_p^2),$$

entonces las variables aleatorias componentes del vector, X_i, $i = 1, \ldots, p$, son independientes y tienen DN univariante (no degenerada),

$$X_i \sim N(\mu_i, \sigma_i^2) \quad (\sigma_i^2 > 0), \quad i = 1, \ldots, p \qquad \text{(□ Probar)}$$

- El recíproco de este resultado es cierto en el siguiente sentido:

  Si $\mathbf{X} = (X_1, \ldots, X_p)'$ es un vector aleatorio con componentes (mutuamente) independientes, cada una de ellas con DN univariante (no degenerada),

$$X_i \sim N(\mu_i, \sigma_i^2) \quad (\sigma_i^2 > 0), \quad i = 1, \ldots, p,$$

  entonces $\mathbf{X}$ tiene DNM no singular,

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma), \quad \text{con} \quad \Sigma = \text{diag}\,(\sigma_1^2, \ldots, \sigma_p^2) \quad (\Sigma > 0)$$

  (OBSERVACIÓN: El resultado no es válido si solo se supone la incorrelación, o incluso la ‘independencia dos a dos’, entre las componentes del vector)

- El Resultado 1 se puede generalizar al caso en que la matriz Σ es una matriz diagonal por cajas, estableciéndose las conclusiones en términos de los correspondientes subvectores del vector $\mathbf{X}$ (se enunciará, por simplicidad, en el caso de dos subvectores):

- RESULTADO 2: Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio con DNM no singular,

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

Supongamos que las componentes de $\mathbf{X}$ están ordenadas de tal modo que para la partición del vector

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} \end{pmatrix}, \quad \text{con} \quad \mathbf{X}_{(1)} = (X_1, \ldots, X_q)', \mathbf{X}_{(2)} = (X_{q+1}, \ldots, X_p)',$$

se tiene

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix} = \begin{pmatrix} \Sigma_{(11)} & 0 \\ 0 & \Sigma_{(22)} \end{pmatrix}$$

(*i. e.*, $\Sigma = \operatorname{diag}\left(\Sigma_{(11)}, \Sigma_{(22)}\right)$). Entonces, los vectores aleatorios $\mathbf{X}_{(1)}$ y $\mathbf{X}_{(2)}$ son (mutuamente) independientes y tienen cada uno DNM, de dimensiones respectivas q y $p-q$,

$$\mathbf{X}_{(1)} \sim N_q(\boldsymbol{\mu}_{(1)}, \Sigma_{(11)})$$
$$\mathbf{X}_{(2)} \sim N_{p-q}(\boldsymbol{\mu}_{(2)}, \Sigma_{(22)})$$

(□ Probar)

- RESULTADO 3: Sea (como en el Resultado 2)

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0), \quad \text{con el particionamiento}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}$$

Entonces, se tiene que

(1) Los vectores $\mathbf{X}_{(1)}$ y $\mathbf{X}_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mathbf{X}_{(1)}$ son independientes

(2) Dichos vectores se distribuyen según

- $\mathbf{X}_{(1)} \sim N_q(\boldsymbol{\mu}_{(1)}, \Sigma_{(11)})$
- $\mathbf{X}_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\mathbf{X}_{(1)} \sim$
  $N_{p-q}\left(\boldsymbol{\mu}_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\boldsymbol{\mu}_{(1)}, \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}\right)$

(3) La distribución condicionada de $\mathbf{X}_{(2)}$ dado $\mathbf{X}_{(1)} = \mathbf{x}_{(1)}$ es una DNM, de la forma

$$N_{p-q}\left(\boldsymbol{\mu}_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)}), \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}\right)$$

($\square$ Probar)

(En adelante, se denotará $\Sigma_{22\cdot 1} = \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}$)

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

La función característica de la DNM, en el caso no singular ($\Sigma > 0$), puede obtenerse de forma directa a partir de la función de densidad

- **CARACTERIZACIÓN II**:

  Un vector aleatorio p-dimensional $\mathbf{X}$ con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\Sigma > 0$ tiene una DNM no singular, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, si y solo si su función característica tiene la forma

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

  (OBSERVACIÓN: Se verá más adelante que en el caso singular, con Σ semidefinida positiva, la función característica tiene la misma expresión formal)

- Frente al uso directo de la función de densidad (caso no singular), la función característica es muy útil en la derivación de diversos resultados, algunos de los cuales se ven a continuación

- RESULTADO 4:

  Sean

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

  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$(B\Sigma_{\mathbf{X}}B') = q$, es decir, $B\Sigma_{\mathbf{X}}B' > 0$) ($\square$ Probar)

- En particular, del resultado anterior se obtiene el siguiente, sobre normalidad de las distribuciones marginales en una DNM

  RESULTADO 5:

  Sea
  $$\mathbf{X} = (X_1, \ldots, X_p)' \sim N_p(\boldsymbol{\mu}, \Sigma) \quad (\Sigma > 0)$$
  Entonces, para todo subvector
  $$\mathbf{X_r} = (X_{r_1}, \ldots, X_{r_q})', \qquad \mathbf{r} = (r_1, \ldots, r_q)',\ q \leq p,$$
  (con $r_1, \ldots, r_q$ distintos entre sí) se tiene que
  $$\mathbf{X_r} \sim N_q(\boldsymbol{\mu}_\mathbf{r}, \Sigma_\mathbf{r}) \quad (\Sigma_\mathbf{r} > 0)$$
  siendo

  $\boldsymbol{\mu}_\mathbf{r}$ el subvector de $\boldsymbol{\mu}$ correspondiente a $\mathbf{r}$
  $\Sigma_\mathbf{r}$ la submatriz principal de Σ correspondiente a $\mathbf{r}$

(□ Probar)

- **CARACTERIZACIÓN III**:

  Un vector aleatorio p-dimensional $\mathbf{X} = (X_1, \ldots, X_p)'$ tiene una DNM no singular si y solo si toda combinación de la forma

  $$\boldsymbol{\alpha}'\mathbf{X}, \quad \text{con } \boldsymbol{\alpha} \in \mathbb{R}^p - \{\mathbf{0}\},$$

  tiene una DN univariante no degenerada (*i. e.*, con varianza no nula)

  (□ Probar)

Caracterización de la DNM en términos de la densidad esférica estándar (caso $\Sigma > 0$)

- Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio con DNM no singular,

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

Para alguna elección de una matriz C no singular con $\Sigma = CC'$, consideramos la normalización de $\mathbf{X}$ dada por el vector aleatorio

$$\mathbf{Z} := C^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, I_p)$$

Definimos ahora el vector aleatorio

$$\mathbf{U} := \frac{\mathbf{Z}}{\|\mathbf{Z}\|} = \frac{\mathbf{Z}}{(\mathbf{Z}'\mathbf{Z})^{\frac{1}{2}}}$$

Puesto que $\|U\| = \left\| \frac{\mathbf{Z}}{\|\mathbf{Z}\|} \right\| = \frac{\|\mathbf{Z}\|}{\|\mathbf{Z}\|} = 1$, se tiene que $\mathbf{U}$ se distribuye sobre la esfera unidad en $\mathbb{R}^p$, $\mathcal{S}_p$. Además, se cumple que

$$\mathbf{U} \stackrel{d}{=} H\mathbf{U}$$

para toda matriz ortogonal H de dimensión p (*i. e.* , $HH' = H'H = I_p$); es decir, $\mathbf{U}$ ha de tener la distribución uniforme sobre $\mathcal{S}_p$ (*distribución esférica* sobre la esfera unidad) (□ Probar)

Caracterización de la DNM en términos de la densidad esférica estándar (caso $\Sigma > 0$) (cont.)

- Un RESULTADO importante:

  El vector aleatorio

  $$\mathbf{U} = \frac{\mathbf{Z}}{(\mathbf{Z}'\mathbf{Z})^{\frac{1}{2}}}$$

  y la variable aleatoria

  $$R = (\mathbf{Z}'\mathbf{Z})^{\frac{1}{2}}$$

  son independientes

  (Este resultado se demuestra, en general, para vectores aleatorios con distribución esférica)

- **CARACTERIZACIÓN IV**:

  Un vector aleatorio p-dimensional $\mathbf{X} = (X_1, \ldots, X_p)'$ tiene una DNM no singular,

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

  si y solo si las realizaciones de $\mathbf{X}$ pueden generarse a partir del siguiente procedimiento (experimento aleatorio):

  1. Observar el valor de una variable aleatoria $R^2 := V \sim \chi^2_p$

     $$\rightarrow \quad \text{Se identifica } V = v$$

  2. Elegir aleatoriamente (*i. e.*, según la distribución uniforme), y de forma independiente, un punto en la esfera p-dimensional de radio $r = \sqrt{v}$

     $$\rightarrow \quad \text{Se identifica como } \mathbf{Z} = (Z_1, \ldots, Z_p)' = (z_1, \ldots, z_p)' = \mathbf{z}$$

  3. Dados un vector de medias $\boldsymbol{\mu}$ y una matriz de covarianzas $\Sigma > 0$, eligiendo C no singular tal que $\Sigma = CC'$, obtener el valor de $\mathbf{X}$ mediante la expresión $\mathbf{X} := C\mathbf{Z} + \boldsymbol{\mu}$

     $$\rightarrow \quad \text{Se identifica } \mathbf{X} = \mathbf{x} = C\mathbf{z} + \boldsymbol{\mu}$$