

# Inferencia Multivariada

Daniel Czarniewicz

## Vectores aleatorios

Una variable aleatoria vectorial (o vector aleatorio) es el resultado de observar  $p$  características de un elemento de una población. Para definir la distribución conjunta de la variable se debe:

- Especificar el espacio muestral,  $\Omega \in \mathbb{R}^p$ .
- Especificar las probabilidades de cada resultado del espacio muestral.

Diremos que el vector aleatorio es discreto si todas sus componentes son discretas. Análogamente, diremos que es continuo si todas sus componentes son continuas. Cuando un vector tenga componentes discretos y continuos, diremos que el mismo es mixto.

## Distribución conjunta

La función de distribución conjunta de un vector aleatorio  $\mathbf{X}$  en un punto  $\mathbf{x}$  se define como:

$$F_{\mathbf{X}}(\mathbf{x}) = \Pr(\mathbf{X} \leq \mathbf{x}) = \Pr(X_1 \leq x_1, \dots, X_p \leq x_p)$$

donde  $F(\cdot)$  es no decreciente.

Llamaremos función de probabilidad de un vector aleatorio discreto a la función  $p(\mathbf{X})$  definida por:

$$p_{\mathbf{X}}(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x}) = \Pr(X_1 = x_1, \dots, X_p = x_p)$$

Diremos que el vector  $\mathbf{X}$  es absolutamente continuo si existe una función de densidad,  $f_{\mathbf{X}}(\mathbf{x})$ , que satisface:

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{X} = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} f_{X_1, \dots, X_p}(x_1, \dots, x_p) dX_1 \dots dX_p$$

donde:

- $f_{\mathbf{X}}(\mathbf{x})$  es no negativa
- $\int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{X} = 1$

La probabilidad de cualquier suceso  $A$  se calcula integrando la función de densidad sobre el subconjunto definido por el suceso.

$$\Pr(A) = \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{X}$$

## Distribución marginal y distribución condicional

Dado un vector aleatorio  $p$ -dimensional, se define la distribución marginal de  $X_i$  como la distribución univariada de dicho componente, es decir:

$$f_{X_i}(x_i) = \int_{\text{Rec}(\mathbf{X}_{-i})} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{X}_{-i} = \int_{\text{Rec}(X_1)} \dots \int_{\text{Rec}(X_{i-1})} \int_{\text{Rec}(X_{i+1})} \dots \int_{\text{Rec}(X_p)} f_{\mathbf{X}}(\mathbf{x}) dX_1 \dots dX_{i-1} dX_{i+1} \dots dX_p$$

De igual forma, puede definirse una partición del vector  $\mathbf{X}$ , por ejemplo,  $\mathbf{X} = (\mathbf{X}_1; \mathbf{X}_2)$ , y definir la distribución marginal conjunta como:

$$f_{\mathbf{X}_i}(\mathbf{x}_i) = \int_{\text{Rec}(\mathbf{X}_j)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{X}_j \quad \text{para } i, j = 1; 2 \quad \text{con } i \neq j$$

Operando de igual forma, se define la distribución condicional de  $\mathbf{X}_1$  a  $\mathbf{X}_2$  como:

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}_2}(\mathbf{x}_2)} = \frac{f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_2}(\mathbf{x}_2)}$$

De lo anterior se desprende que la distribución marginal de  $\mathbf{X}_i$  puede plantearse como:

$$f_{\mathbf{X}_i}(\mathbf{x}_i) = \int_{\text{Rec}(\mathbf{X}_j)} f_{\mathbf{X}_i|\mathbf{X}_j}(\mathbf{x}_i|\mathbf{x}_j) f_{\mathbf{X}_j}(\mathbf{x}_j) d\mathbf{X}_j$$

## Teorema de Bayes

Podemos entonces explicitar el teorema de Bayes para funciones de densidad como:

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) f_{\mathbf{X}_1}(\mathbf{x}_1)}{\int_{\text{Rec}(\mathbf{X}_1)} f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) f_{\mathbf{X}_1}(\mathbf{x}_1) d\mathbf{X}_1}$$

O, en forma general, como:

$$f_{\mathbf{X}_i|\mathbf{X}_j}(\mathbf{x}_i|\mathbf{x}_j) = \frac{f_{\mathbf{X}_j|\mathbf{X}_i}(\mathbf{x}_j|\mathbf{x}_i) f_{\mathbf{X}_i}(\mathbf{x}_i)}{\int_{\text{Rec}(\mathbf{X}_i)} f_{\mathbf{X}_j|\mathbf{X}_i}(\mathbf{x}_j|\mathbf{x}_i) f_{\mathbf{X}_i}(\mathbf{x}_i) d\mathbf{X}_i}$$

En todo lo anterior, para el caso de vectores discretos, simplemente se deben cambiar las integrales por sumas.

## Independencia

Diremos que dos vectores aleatorios  $\mathbf{X}_1$  y  $\mathbf{X}_2$  son independientes si el conocimiento de uno de ellos no aporta información respecto de los valores del otro. Matemáticamente:

$$f(\mathbf{x}_1|\mathbf{x}_2) = f(\mathbf{x}_1)$$

Por lo tanto:

$$f(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1) f(\mathbf{x}_2)$$

Si dos vectores  $p$ -dimensionales son independientes, también los son subconjuntos de dimensión  $h < p$  de variables constituyentes, y funciones de las mismas.

## Propiedades de los vectores aleatorios

### Vector de medias

La esperanza de un vector aleatorio  $p$ -dimensional se define simplemente como un vector de largo  $p$ , donde cada elemento corresponde a la esperanza de una componente de  $\mathbf{X}$ .

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = \int_{\mathbf{X}} \mathbf{x} f(\mathbf{x}) d\mathbf{X}$$

La esperanza es un operador lineal, al igual que en el caso univariado. Esto implica que:

- para toda matriz  $\mathbf{A}$  y vector  $\mathbf{b}$ ,  $\mathbb{E}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\mathbb{E}(\mathbf{X}) + \mathbf{b}$
- para todo grupo de vectores aleatorios,  $\sum_{i=1}^n \mathbb{E}(\mathbf{X}_i) = \mathbb{E}\left(\sum_{i=1}^n \mathbf{X}_i\right)$
- para todo par de vectores independientes,  $\mathbb{E}(\mathbf{XY}) = \mathbb{E}(\mathbf{X}) \mathbb{E}(\mathbf{Y})$

Análogamente con el caso univariado, la esperanza de una función  $g$  del vector aleatorio será:

$$\mathbb{E}[g(\mathbf{X})] = \int_{\mathbf{X}} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{X}$$

### Matriz de varianzas y covarianzas

Llamaremos matriz de varianzas y covarianzas de un vector aleatorio  $\mathbf{X}$  de dimensión  $p$  a la matriz cuadrada de orden  $p$  obtenida por:

$$\mathbf{V}_x = \mathbb{V}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\right]_{p \times p} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}_{p \times p}$$

donde  $\sigma_i^2$  representa la varianza de la  $i$ -ésima componente, y  $\sigma_{ij}$  la covarianza entre la componente  $i$  y la componente  $j$ . La matriz es semidefinida positiva, es decir, dado un vector  $\mathbf{a}$  de  $\mathbb{R}^p$ ,  $\mathbf{a}'\mathbf{V}_x\mathbf{a} \geq 0$ .

Medidas de variabilidad:

- Varianza media:  $VM = \text{tr}(\mathbf{V}_x)/p$
- Varianza generalizada:  $VG = \det(\mathbf{V}_x)$
- Varianza promedio:  $VP = \left(\det(\mathbf{V}_x)\right)^{1/p}$

## Transformaciones

Sea  $\mathbf{X}$  un vector aleatorio de  $\mathbb{R}^p$  con función de densidad  $f_{\mathbf{X}}(\mathbf{x})$ , y otro vector aleatorio  $\mathbf{Y}$  de  $\mathbb{R}^p$  definido mediante la transformación uno a uno:

$$\begin{aligned} y_1 &= g_1(\mathbf{X}) \\ \vdots &\quad \quad \vdots \\ y_p &= g_p(\mathbf{X}) \end{aligned}$$

tal que existen las funciones inversas  $x_1 = h_1(\mathbf{Y}), \dots, x_p = h_p(\mathbf{Y})$ , y todas ellas son diferenciables. Entonces, la función de densidad del vector aleatorio  $\mathbf{Y}$  estará dada por:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(h(\mathbf{Y})) \left| \frac{d\mathbf{X}}{d\mathbf{Y}} \right|$$

donde el segundo factor es el valor absoluto del Jacobiano de la matriz (es decir, el valor absoluto del determinante de la matriz Jacobiana):

$$\left| \frac{d\mathbf{X}}{d\mathbf{Y}} \right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_p}{\partial y_1} & \cdots & \frac{\partial x_p}{\partial y_p} \end{vmatrix}$$

Ante la transformación lineal  $\mathbf{Y} = \mathbf{A}\mathbf{X}$ , donde  $\mathbf{A}$  es una matriz cuadrada no singular:

- $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{Y})|\mathbf{A}|^{-1}$

Si, en cambio,  $\mathbf{X}$  es de dimensión  $p$  mientras que  $\mathbf{Y}$  es de dimensión  $m \leq p$ , entonces:

- $\mathbf{Y} = \mathbf{A}\mathbf{X}$  donde  $\mathbf{A} \in \mathcal{M}_{m \times p}$
- $\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}}$
- $\mathbf{V}_{\mathbf{Y}} = \mathbf{A}\mathbf{V}_{\mathbf{X}}\mathbf{A}'$

## Matriz de correlaciones

Se define la matriz de correlaciones de un vector aleatorio  $\mathbf{X}$  con matriz de varianzas  $\mathbf{V}_x$  como:

$$\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{V}_x \mathbf{D}^{-1/2}$$

donde  $\mathbf{D} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ . Por lo tanto, la matriz de correlaciones será una matriz cuadrada, simétrica, semidefinida positiva, y con unos en la diagonal principal. Los elementos fuera de la diagonal representan la correlación entre los márgenes del vector aleatorio,  $\rho_{ij} = \sigma_{ij} / \sigma_i \sigma_j$ .

Una medida global de la dependencia viene definida por:

$$D_x = 1 - |\mathbf{R}|^{1/(p-1)}$$

## Maldición de la alta dimensionalidad

Este término es utilizado para describir como aumenta la complejidad de un problema al aumentar la dimensión (cantidad de variables involucradas). En el análisis estadístico se manifiesta de dos formas.

Primero, al aumentar la dimensión del problema, el espacio está cada vez más vacío, haciendo más difícil cualquier proceso de inferencia a partir de los datos. Esto se debe a que la masa de probabilidad sigue siendo uno a pesar del aumento de la dimensionalidad del problema.

Un segundo problema es el aumento del número de parámetros que se necesita estimar para describir los datos. Por ejemplo, si se desean estimar la media y la matriz de varianzas de un vector  $p$ -dimensional, se requieren estimar  $p + p(p+1)/2 = p(p+3)/2$  parámetros, lo cual es de orden  $p^2$ . Es decir, la cantidad de parámetros a estimar aumenta con el cuadrado de la dimensión. Esto hace que se necesiten cada vez más datos para estimar problemas de alta dimensionalidad. Como norma general, se desea que la cantidad de observaciones sea del orden de  $n/p > 10$ .

En resumen, el problema del aumento de la dimensionalidad es el aumento de la incertidumbre del problema. Este disminuye si existe alta dependencia entre las variables, dado que la masa de probabilidad se concentra en zonas del espacio. Dichas zonas son definidas por la relación de dependencia.

## Distribución Normal Multivariada

Un vector aleatorio  $\mathbf{X}$  tiene distribución normal multivariada si su función de densidad es:

$$f(\mathbf{x}) = |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

y escribiremos que  $\mathbf{X} \sim N_p(\boldsymbol{\mu}_x, \mathbf{V}_x)$ .

## Propiedades

1. La distribución es simétrica respecto de  $\boldsymbol{\mu}$ .

$$\begin{aligned}
 f(\boldsymbol{\mu} + \mathbf{a}) &= |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} + \mathbf{a} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\boldsymbol{\mu} + \mathbf{a} - \boldsymbol{\mu}) \right\} \\
 &= |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} \mathbf{a}' \mathbf{V}_x^{-1} \mathbf{a} \right\} \\
 f(\boldsymbol{\mu} - \mathbf{a}) &= |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \mathbf{a} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\boldsymbol{\mu} - \mathbf{a} - \boldsymbol{\mu}) \right\} \\
 &= |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (-\mathbf{a})' \mathbf{V}_x^{-1} (-\mathbf{a}) \right\} \\
 &= |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} \mathbf{a}' \mathbf{V}_x^{-1} \mathbf{a} \right\}
 \end{aligned}$$

2. La distribución tiene un único máximo en  $\boldsymbol{\mu}$ .

Dado que  $\mathbf{V}_x$  es definida positiva, el término del exponente es siempre positivo, por lo que la función tendrá un máximo cuando este sea cero, lo que se cumple cuando  $\mathbf{x} = \boldsymbol{\mu}$ .

3. La esperanza de  $\mathbf{X}$  es  $\boldsymbol{\mu}$ .
4. La varianza de  $\mathbf{X}$  es  $\mathbf{V}_x$ .
5. Si la  $p$  variables tienen distribución conjunta normal multivariada y están incorreladas, entonces son independientes.
6. Si  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ , entonces, sus marginales  $X_i \sim N(\mu_i, \sigma_i^2)$ .
7. Cualquier subconjunto de las marginales de una normal, distribuye normal.
8. Sea  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) \sim N_p(\boldsymbol{\mu}, \mathbf{V})$  donde:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{y} \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$$

Entonces,

$$\begin{aligned}
 \mathbf{X}_2 | \mathbf{X}_1 &\sim N_m \left( \boldsymbol{\mu}_2 + \mathbf{V}_{12}' \mathbf{V}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \mathbf{V}_{22} - \mathbf{V}_{12}' \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \right) \\
 \mathbf{X}_1 | \mathbf{X}_2 &\sim N_n \left( \boldsymbol{\mu}_1 + \mathbf{V}_{22}^{-1} \mathbf{V}_{21}' (\mathbf{x}_2 - \boldsymbol{\mu}_2), \mathbf{V}_{11} - \mathbf{V}_{21} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}' \right)
 \end{aligned}$$

## Inferencia Multivariada

### Estimación máximo verosimil

El método de MV escoge como estimadores de los parámetros aquellos valores que hacen máxima la probabilidad de observar la muestra obtenida. Supongamos una MAS de  $\mathbf{X} \sim f(\mathbf{x}|\boldsymbol{\theta})$  con  $\boldsymbol{\theta} \in \mathbb{R}^r$ , donde  $r \leq pn$ . La función de densidad conjunta de la muestra será:

$$f(\mathbf{X}|\boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{x}_i|\boldsymbol{\theta})$$

Para construir la función de verosimilitud simplemente debe mirarse dicha función como función de  $\boldsymbol{\theta}$  dado  $\mathbf{x}$ :

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = \prod_{i=1}^n f(\mathbf{x}_i|\boldsymbol{\theta})$$

El estimador MV será el resultado de resolver:

$$\frac{\partial}{\partial \theta_1} \mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = 0, \dots, \frac{\partial}{\partial \theta_r} \mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = 0$$

con Hessiana definida negativa (condición de segundo orden del máximo).

Suele ser más sencillo utilizar la log-verosimilitud, definida como:

$$\ell(\boldsymbol{\theta}|\mathbf{X}) = \log(\mathcal{L}(\boldsymbol{\theta}|\mathbf{X}))$$

### Estimación MV para normal multivariada

Para el caso de un vector aleatorio normal multivariado la función de densidad conjunta de la muestra será:

$$f(\mathbf{X}|\boldsymbol{\mu}, \mathbf{V}) = \prod_{i=1}^n |\mathbf{V}_x|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Su log-verosimilitud estará dado por:

$$\ell(\boldsymbol{\mu}, \mathbf{V}|\mathbf{X}) = -\frac{n}{2} \log |\mathbf{V}_x| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

Sea  $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ , y, teniendo en cuenta que  $(\mathbf{x}_i - \boldsymbol{\mu}) = (\mathbf{x}_i - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \boldsymbol{\mu})$  obtenemos que:

$$\sum_{i=1}^n (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \sum_{i=1}^n (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{V}_x^{-1} (\mathbf{x} - \bar{\mathbf{x}}) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

Dado que un escalar es igual a su traza:

$$\text{tr} \left( \sum_{i=1}^n (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{V}_x^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \right) = \text{tr} \left( \mathbf{V}_x^{-1} \sum_{i=1}^n (\mathbf{x} - \bar{\mathbf{x}})' (\mathbf{x} - \bar{\mathbf{x}}) \right) = \text{tr} \left( \mathbf{V}_x^{-1} n \mathbf{S} \right)$$

Reemplazando en la función de log-verosimilitud obtenemos que:

$$\ell(\boldsymbol{\mu}, \mathbf{V} | \mathbf{X}) = -\frac{n}{2} \log |\mathbf{V}_x| - \frac{1}{2} \text{tr} \left( \mathbf{V}_x^{-1} n \mathbf{S} \right) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

$\mathbf{V}_x^{-1}$  es definida positiva  $\Rightarrow (\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \geq 0 \Rightarrow -\frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq 0$ , por lo que la expresión se maximiza cuando

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$$

Reemplazando  $\boldsymbol{\mu}$  por su estimación MV en la función de log-verosimilitud, y sumando la constante  $\frac{n}{2} \log |\mathbf{S}|$ , obtenemos que:

$$\ell(\mathbf{V} | \mathbf{X}) = \frac{n}{2} \log |\mathbf{V}_x^{-1} \mathbf{S}| - \frac{n}{2} \text{tr} \left( \mathbf{V}_x^{-1} \mathbf{S} \right)$$

Sean  $\lambda_i$  los valores propios de  $\mathbf{V}_x^{-1} \mathbf{S}$ , entonces:

$$\ell(\mathbf{V} | \mathbf{X}) = \frac{n}{2} \sum_i \log \lambda_i - \frac{n}{2} \sum_i \lambda_i = \frac{n}{2} \sum_i [\log \lambda_i - \lambda_i]$$

Maximizando respecto de cada  $\lambda_i$  obtenemos que:

$$\frac{\partial \ell(.)}{\partial \lambda_i} = \frac{n}{2} \sum_i \left[ \frac{1}{\lambda_i} - 1 \right] = 0 \Rightarrow \frac{1}{\lambda_i} - 1 = 0 \Rightarrow \lambda_i = 1 \quad \forall i$$

Lo cual implica que  $\mathbf{V}_x^{-1} \mathbf{S} = \mathbf{I}$ , por lo que:

$$\hat{\mathbf{V}}_x = \mathbf{S}$$

Algunas consideraciones finales:

1.  $\ell(.)$  solo depende de  $\mathbf{x}$  a través de  $\bar{\mathbf{x}}$  y  $\mathbf{S}$ , por lo que  $(\bar{\mathbf{x}}, \mathbf{S})$  es suficiente para  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{V}_x)$
2. Si  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \mathbf{V}_x)$ , entonces  $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \mathbf{V}_x/n)$ .
3.  $\mathbf{S}$  es sesgado, pero  $\frac{n}{n-1} \mathbf{S}$  es insesgado para  $\mathbf{V}_x$ .
4.  $\mathbf{S}$  y  $\bar{\mathbf{x}}$  son consistentes, eficientes, y asintóticamente normales.



## Contraste sobre la media

Consideremos una muestra proveniente de una normal multivariada de media  $\boldsymbol{\mu}$ , y matriz de covarianzas  $\mathbf{V}_x$ . Se desea contrastar la hipótesis:

$$\begin{aligned} H_0 : \boldsymbol{\mu} &= \boldsymbol{\mu}_0, \quad \mathbf{V}_x \text{ cualquiera} \\ H_1 : \boldsymbol{\mu} &\neq \boldsymbol{\mu}_0, \quad \mathbf{V}_x \text{ cualquiera} \end{aligned}$$

La log-verosimilitud será:

$$\ell(\boldsymbol{\mu} | \mathbf{V}_x, \mathbf{X}) = -\frac{n}{2} \log |\mathbf{V}_x| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

Bajo  $H_1$ ,  $\hat{\boldsymbol{\mu}}_{MV} = \bar{\mathbf{x}}$  y  $\hat{\mathbf{V}}_{MV} = \mathbf{S}$ , por lo que

$$\ell(H_1) = -\frac{n}{2} \log |\mathbf{S}| - \frac{np}{2}$$

Mientras que bajo  $H_0$ ,  $\hat{\boldsymbol{\mu}}_{MV} = \boldsymbol{\mu}_0$  y  $\hat{\mathbf{V}}_{MV} = \mathbf{S}_0 = \frac{1}{n} \sum_{i=1}^n (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'$ , por lo que

$$\ell(H_0) = -\frac{n}{2} \log |\mathbf{S}_0| - \frac{np}{2}$$

La diferencia será entonces:

$$\lambda = -2 \ln(RV) = 2[\ell(H_1) - \ell(H_0)] \Rightarrow \lambda = n \log \left( \frac{|\mathbf{S}_0|}{|\mathbf{S}|} \right) \stackrel{a}{\sim} \chi_p^2$$

Los grados de libertad del estadístico corresponden a la diferencia en las dimensiones en que se mueven los parámetros bajo  $H_0$  y bajo  $H_1$ :

- Bajo  $H_0$ :  $p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2}$
- Bajo  $H_1$ :  $p + \frac{p(p+1)}{2} = \frac{p(p+3)}{2}$

Si la cantidad de observaciones no es lo suficientemente alto, y por tanto no puede utilizarse una distribución asintótica para el estimador, entonces se utilizar la distribución  $T^2$  de Hotelling donde:

- si  $p = 1 \Rightarrow T^2 \rightarrow t_{n-1}$
- si  $n \rightarrow +\infty \Rightarrow T^2 \rightarrow \chi_p^2$

## Contraste sobre la matrix de covarianzas

### Contraste de valor particular

$$\begin{aligned} H_0 : \mathbf{V} &= \mathbf{V}_0, \quad \boldsymbol{\mu} \text{ cualquiera} \\ H_1 : \mathbf{V} &\neq \mathbf{V}_0, \quad \boldsymbol{\mu} \text{ cualquiera} \end{aligned}$$

Log-verosimilitud:

$$\ell(\mathbf{V}|\boldsymbol{\mu}\mathbf{X}) = -\frac{n}{2} \log |\mathbf{V}| - \frac{n}{2} \text{tr}(\mathbf{V}^{-1}\mathbf{S}) - \frac{n}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

Bajo  $H_0$ :  $\hat{\mathbf{V}} = \mathbf{V}_0$  y  $\hat{\boldsymbol{\mu}}_{MV} = \bar{\mathbf{x}}$  por lo que

$$\ell(H_0) = -\frac{n}{2} \log |\mathbf{V}_0| - \frac{n}{2} \text{tr}(\mathbf{V}_0^{-1}\mathbf{S})$$

Bajo  $H_0$ :  $\hat{\mathbf{V}} = \mathbf{S}$  y  $\hat{\boldsymbol{\mu}}_{MV} = \bar{\mathbf{x}}$  por lo que

$$\ell(H_1) = -\frac{n}{2} \log |\mathbf{S}| - \frac{np}{2}$$

Estadístico RV:

$$\lambda = n \log \left( \frac{|\mathbf{V}_0|}{|\mathbf{S}|} \right) + n \text{tr}(\mathbf{V}_0^{-1}\mathbf{S}) - np \stackrel{a}{\sim} \chi^2_{\frac{p(p+1)}{2}}$$

Si  $\mathbf{V}_0 = \mathbf{I} \Rightarrow \lambda = -n \log |\mathbf{S}| + n \text{tr}(\mathbf{S}) - np$

### Contraste de independencia

$$\begin{aligned} H_0 : \mathbf{V} &\text{ diagonal, } \boldsymbol{\mu} \text{ cualquiera} \\ H_1 : \mathbf{V} &\text{ y } \boldsymbol{\mu} \text{ cualesquiera} \end{aligned}$$

Bajo  $H_0$ ,  $\mathbf{V} = \mathbf{V}_0$ , y  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ .

Bajo  $H_1$ ,  $\hat{\mathbf{V}}_0 = \text{diag}(\mathbf{S})$ , donde  $\text{diag}(\mathbf{S})$  tiene términos  $s_{ii}$ .

El estadístico RV está dado por:

$$\lambda = n \log \left[ \frac{\prod_{i=1}^n s_{ii}}{|\mathbf{S}|} \right] + n \text{tr}(\hat{\mathbf{V}}_0^{-1}\mathbf{S}) - np = -n \sum_{i=1}^n \log \lambda_i \stackrel{a}{\sim} \chi^2_{\frac{p(p-1)}{2}}$$

donde  $\lambda_i$  son los valores propios de la matrix  $\mathbf{R} = \hat{\mathbf{V}}_0^{-1/2} \mathbf{S} \hat{\mathbf{V}}_0^{-1/2}$ .

**Contraste de esfericidad**

$$H_0 : \mathbf{V} = \sigma^2 \mathbf{I}, \boldsymbol{\mu} \text{ cualquiera}$$

$$H_1 : \mathbf{V} \text{ y } \boldsymbol{\mu} \text{ cualesquiera}$$

Bajo  $H_0$ :

$$\ell(H_0) = -\frac{np}{2} \log(\sigma^2) - \frac{n}{2\sigma^2} \text{tr}(\mathbf{S})$$

$$\frac{\partial \ell(H_0)}{\partial \sigma^2} = -\frac{np}{2} \frac{1}{\sigma^2} + \frac{n \text{tr}(\mathbf{S})}{2\sigma^4} = 0 \Rightarrow \hat{\sigma}^2 = \frac{\text{tr}(\mathbf{S})}{p}$$

Bajo  $H_1$ ,  $\hat{\mathbf{V}}_0 = \text{diag}(\mathbf{S})$  con  $s_{ii} = [\mathbf{S}]_{ii}$

Estadístico RV:

$$\begin{aligned} \lambda &= n \log \left[ \frac{\hat{\sigma}^2}{|\mathbf{S}|} \right] + n \frac{\text{tr}(\mathbf{S})}{\hat{\sigma}^2} - np \\ &= np \log(\hat{\sigma}^2) - n \log |\mathbf{S}| \sim \chi^2_{\frac{(p+2)(p-1)}{2}} \end{aligned}$$

**Test de homogeneidad**

Aquí se detalla el test de Box-M, una versión multivariada del test de Bartlett. El objetivo es testear si existe evidencia de que las matrices de covarianza de las distintas subpoblaciones en la muestra sean distintas.

$$H_0 : \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_G$$

$$H_1 : \exists \boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j$$

Estadístico:

$$-2(1 - c_1) \ln M \sim \chi_v^2$$

con:

$$M = \frac{1}{2} \sum_{g=1}^G (n_g - 1) \log |\mathbf{S}_g| - \frac{1}{2} (n - k) \log |\mathbf{S}|$$

donde:

- $\mathbf{S}$  es la pooled covariance matrix

$$\bullet \quad c_1 = \left[ \sum_{k=1}^G \frac{1}{n_k - 1} - \frac{1}{\sum_{k=1}^G n_k - 1} \right] \left[ \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right]$$

$$\bullet \quad v = \frac{1}{2} (k-1)p(p+1)$$

## Test de normalidad

La normalidad multivariante implica la normalidad de las marginales, pero la normalidad de las marginales no implica la normalidad multivariada. Por lo tanto, es importante realizar tests multivariados de normalidad. Existen versiones multivariadas de todos los tests clásicos. Aquí se presenta el test de Mardia, el cual se basa en comparar asimetría y curtosis.

Sean el coeficiente de asimetría  $A_p$  y de curtosis  $K_p$  tales que:

$$A_p = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^3$$

$$K_p = \frac{1}{n} \sum_{i=1}^n d_{ii}^2$$

donde  $d_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$ . Asintóticamente tenemos que:

$$\frac{n A_p}{6} \sim \chi_f^2 \quad \text{con} \quad f = \frac{p(p+1)(p+2)}{6}$$

$$K_p \sim N \left( p(p+2), \frac{8p(p+2)}{n} \right)$$

## Referencias

Blanco, Jorge, and others. 2006. “Introducción Al análisis Multivariado.” *IESTA. Montevideo*.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

Peña, Daniel. 2013. *Análisis de Datos Multivariantes*. McGraw-Hill España.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Rencher, Alvin C. 1998. *Multivariate Statistical Inference and Applications*. Wiley New York.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.