

# MGE-201: Modelo Estadístico

**Felipe Osorio**

[felipe.osorio@uv.cl](mailto:felipe.osorio@uv.cl)

## Datos:

Consideraremos que las **observaciones** corresponden a números, vectores, categorías o funciones. En general, diremos que

$$\mathbf{y} \in \mathcal{Y},$$

con  $\mathcal{Y}$  el **espacio muestral**.

## Observación:

Asumiremos que  $\mathbf{y}$  es el resultado de un **experimento aleatorio**. Es decir, la realización de una variable aleatoria  $Y$ .

## Idea:

El objetivo de la **inferencia estadística** es obtener información sobre la distribución de  $Y$  a partir de los datos  $\mathbf{y}$ .

De este modo, asumiremos que  $Y$  es un miembro de la familia

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\},$$

que es indexada por  $\theta \in \Theta$ . El conjunto  $\Theta$  es denominado **espacio paramétrico**.

## Idea:

Es decir, se desea conocer algunas propiedades de  $\mathcal{P}$  basado en una (única) muestra aleatoria.

## Definición 1 (modelo estadístico):

Un **modelo estadístico** es un par  $(\mathcal{Y}, \mathcal{P})$ , donde  $\mathcal{Y}$  denota el **espacio muestral**.

## Definición 2 (muestra aleatoria):

Una **muestra aleatoria**<sup>1</sup>  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  es una colección de variables aleatorias independientes, donde cada  $Y_i$  es distribuída como  $P_{i,\theta}$ .<sup>2</sup>

### *Observación:*

Una **muestra IID**  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  es una colección donde las variables aleatorias son independientes y tienen la **misma distribución**. Es decir,

$$P_{i,\theta} = P_\theta, \quad \forall i.$$

---

<sup>1</sup>Usualmente utilizaremos la notación m.a.( $n$ ).

<sup>2</sup>El **tamaño muestral**  $n$  es el número de variables aleatorias.

Cada  $Y_j$  tiene distribución acumulada **común**  $F$ . Si  $F$  es **conocido**, podemos usar cálculo de probabilidades para deducir y estudiar sus propiedades.

### *Observación:*

En la práctica,  $F$  es **desconocido**, y el objetivo es tratar de inferir sus propiedades **desde los datos**.

Frecuentemente, el interés es una **función no aleatoria** de  $F$ , tal como la media o su  $q$ -ésimo cuantil.

$$E(Y) = \int y dF(y), \quad y_q = F^{-1}(q) = \inf\{y : F(y) \geq q\}.$$

Las cantidades  $E(Y)$ ,  $\text{var}(Y)$  y  $F^{-1}(q)$  son llamadas **parámetros**, dependen de  $F$  y son generalmente desconocidos.

## Ejemplo:

Considere  $X \sim \text{Bin}(n, \theta)$  es decir

$$p(x; \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x},$$

con  $n \in \mathbb{N}$ ,  $\theta \in (0, 1)$  y  $x \in \{0, 1, \dots, n\}$ . Es decir  $\mathcal{X} = \{0, 1, \dots, n\}$  y  $\Theta = (0, 1)$ .<sup>3</sup>

De este modo,

$$\mathcal{P} = \{\text{Bin}(n, \theta) : \theta \in (0, 1)\}.$$

---

<sup>3</sup>Estamos considerando  $n$  fijo

### Ejemplo:

Suponga que  $\mathbf{X} = (X_1, \dots, X_n)^\top$  son IID tal que  $X_i \sim \text{Poi}(\theta)$ ,  $i = 1, \dots, n$ . Aquí

$$p(x_i; \theta) = \frac{e^{-\theta} \theta^{x_i}}{x_i!}, \quad \theta \in (0, \infty),$$

y  $x_i \in \{0, 1, \dots\}$ . De este modo,<sup>4</sup>

$$\mathcal{P} = \{\text{Poi}(\theta)^{\otimes n} : \theta \in (0, \infty)\}.$$

Note que la **densidad conjunta** (asociada a  $\text{Poi}(\theta)^{\otimes n}$ ) es dada por

$$p(\mathbf{x}; \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

---

<sup>4</sup>Además,  $\mathcal{X} = \{0, 1, \dots\} \times \dots \times \{0, 1, \dots\} = \{0, 1, \dots\}^n$ .

## *Ejemplo:*

Suponga  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  con

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \right\}.$$

En este caso tenemos,  $\mathcal{Y} = \mathbb{R}^p$ , mientras que el modelo estadístico es definido como

$$\mathcal{P} = \{N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} > \mathbf{0}\}.$$

## *Observación:*

Note que el modelo tiene  $p + p(p + 1)/2$  parámetros ( $\boldsymbol{\Sigma}$  es matriz simétrica).



## *Ejemplo (Estructura de equicorrelación):*

Considere  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  vectores aleatorios IID desde  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\phi}))$  con

$$\boldsymbol{\Sigma}(\boldsymbol{\phi}) = \phi_1[\phi_2 \mathbf{I} + (1 - \phi_2) \mathbf{1}\mathbf{1}^\top] = \phi_1 \begin{pmatrix} 1 & \phi_2 & \dots & \phi_2 \\ \phi_2 & 1 & \dots & \phi_2 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_2 & \phi_2 & \dots & 1 \end{pmatrix},$$

donde  $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$ . En este caso tenemos,

$$\begin{aligned} f(\mathbf{Y}; \boldsymbol{\mu}, \boldsymbol{\phi}) &= |2\pi \boldsymbol{\Sigma}(\boldsymbol{\phi})|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}) (\mathbf{y}_i - \boldsymbol{\mu}) \right\} \\ &= |2\pi \boldsymbol{\Sigma}(\boldsymbol{\phi})|^{-n/2} \exp \left\{ -\frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}) \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top \right\} \\ &= |2\pi \boldsymbol{\Sigma}(\boldsymbol{\phi})|^{-n/2} \exp \left\{ -\frac{n}{2} \left[ \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}) \mathbf{S} + (\bar{\mathbf{y}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}) (\bar{\mathbf{y}} - \boldsymbol{\mu}) \right] \right\}, \end{aligned}$$

con  $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top$  y  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)^\top \in \mathbb{R}^{n \times p}$ .

De este modo,

$$\mathcal{P} = \{N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})^{\otimes n} : \boldsymbol{\mu} \in \mathbb{R}^p, \phi_1 > 0, -1/(p-1) < \phi_2 < 1\},$$

o equivalentemente,

$$\mathcal{P} = \{N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})^{\otimes n} : \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma}(\boldsymbol{\phi}) > \mathbf{0}\}.$$

Es decir,  $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top)^\top$  es vector  $p + 2$  dimensional.

*Observación:*

Note que podríamos ‘reparametrizar’ el modelo como:<sup>5</sup>

$$\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \alpha_1 \mathbf{I} + \alpha_2 \mathbf{1}\mathbf{1}^\top, \quad \alpha_1 = \phi_1 \phi_2, \quad \alpha_2 = \phi_1 (1 - \phi_2).$$

---

<sup>5</sup>¿Cómo luce el espacio paramétrico ahora?

### *Ejemplo (Regresión lineal simple):*

Considere el modelo de regresión:

$$Y_t = a + bt + \epsilon_t, \quad t = 1, \dots, n.$$

Además, asumiremos que los errores satisfacen

$$E(\epsilon_t) = 0, \quad \text{var}(\epsilon_t) = \sigma^2,$$

y  $E(\epsilon_t \epsilon_r) = 0$  para  $t \neq r$ . De esta manera el modelo para  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  toma valores en  $\mathcal{Y} = \mathbb{R}^n$  y puede ser escrito como:

$$\begin{aligned} \mathcal{P} &= \{P_\theta : E(Y_t) = a + bt, \quad \text{var}(Y_t) = \sigma^2, \quad 1 \leq t \leq n; \\ &\quad \theta = (a, b, \sigma^2) \in \Theta \subset \mathbb{R}^2 \times \mathbb{R}_+\}, \end{aligned}$$

### *Observación:*

Usualmente se asume<sup>6</sup> que la distribución que caracteriza los primeros momentos de  $\{\epsilon_t\}$  es conocida.

---

<sup>6</sup>por ejemplo en el curso [MAT266: Análisis de Regresión](#)

### Observación:

La parametrización usada no es única. En lugar de  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ , podemos escoger

$$\mathcal{P} = \{P_\phi : \phi \in \Phi\},$$

donde  $\theta = h^{-1}(\phi)$  para alguna función 1 a 1.

### Ejemplo:

Considere  $X \sim \text{Exp}(\lambda)$  con densidad

$$f(x; \lambda) = \lambda \exp(-\lambda x), \quad x \geq 0, \lambda > 0.$$

En este caso,  $E(X) = 1/\lambda = \mu$ . Esto lleva a la parametrización alternativa

$$f(x; \mu) = \frac{1}{\mu} \exp(-x/\mu), \quad x \geq 0, \mu > 0.$$

### *Ejemplo:*

Considere  $X \sim \text{Poisson}(\mu)$  con función de probabilidad

$$p(x; \mu) = \frac{\mu^x e^{-\mu}}{x!}, \quad \mu > 0,$$

de ahí que

$$\log p(x; \mu) = x \log \mu - \mu - \log x!$$

Podemos considerar  $\theta = \log \mu$ ,

$$\log p(x; \theta) = x\theta - e^\theta - \log x!, \quad \theta \in \mathbb{R}.$$

Que puede ser de bastante utilidad cuando enfrentemos el problema de estimación de parámetros.

## Supuesto A0 (Identificabilidad):

Para  $\theta_1, \theta_2 \in \Theta$  con  $\theta_1 \neq \theta_2$ . Si las distribuciones  $P_{\theta_1}$  y  $P_{\theta_2}$  son diferentes, entonces se dice que el modelo es **identificable**.

## Definición 3 (Estadística):

Una **estadística**  $T$  es una función de la muestra. Es decir,

$$T : \mathcal{Y} \rightarrow \mathcal{T},$$

tal que  $T(\mathbf{y}) = t$ .

Note además que  $T(\mathbf{Y})$  es una variable aleatoria con función de densidad  $f(t; \theta)$ .

### Ejemplo:

Considere, las siguientes estadísticas

$$T_1(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X},$$

$$T_2(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2.$$

Tenemos  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^2$ , con  $\mathbf{T}(X_1, \dots, X_n) \rightarrow (T_1, T_2)$ .

### Ejemplo:

Considere la función de **distribución empírica**:

$$\hat{F}(X_1, \dots, X_n)(x) = \hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

donde  $(X_1, \dots, X_n)^\top$  es una muestra desde la distribución  $F$  y  $I(\cdot)$  representa la función indicatriz.

### Observación:

Un estadístico depende **sólo** de la muestra. **No** puede depender de cantidades desconocidas. Por ejemplo, para poblaciones normales

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i (= T_1), \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 (= T_2),$$

son **estadísticas** para  $\mu$  y  $\sigma^2$ , respectivamente. Mientras que

$$T_3 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2,$$

**no** es un estadístico para  $\sigma^2$  pues depende de  $\mu$ .