MAT-206: Modelo Estadístico

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Datos:

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

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

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

Observación:

Asumiremos que $\boldsymbol{y}$ 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 $\boldsymbol{y}$.

Modelo estadístico

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

$$\mathcal{P} = \{\mathsf{P}_\theta : \boldsymbol{\theta} \in \Theta\},$$

que es indexada por $\boldsymbol{\theta} \in \Theta$. El conjunto Θ 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[1] $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$ es una colección de variables aleatorias independientes, donde cada Y_i es distribuída como $\mathsf{P}_{i,\theta}$.[2]

Observación:

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

$$\mathsf{P}_{i,\theta} = \mathsf{P}_\theta, \quad \forall i.$$

[1] Usualmente utilizaremos la notación m.a.(n).

[2] El tamaño muestral n es el número de variables aleatorias.

Modelo estadístico

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.

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

Las cantidades $\mathsf{E}(Y)$, $\mathsf{var}(Y)$ 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 $x \in \{0, 1, \ldots, n\}$. Es decir $\mathcal{X} = \{0, 1, \ldots, n\}$ y $\Theta = (0, 1)$.[3]
De este modo,

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

[3] Estamos considerando n fijo

Ejemplo:

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

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

y $x_i \in \{0, 1, \dots\}$. De este modo,[4]

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

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

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

[4] Además, $\mathcal{X} = \{0, 1, \dots\} \times \cdots \times \{0, 1, \dots\} = \{0, 1, \dots\}^n$.

Modelo estadístico

Ejemplo:

Suponga $\boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con

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

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

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

Observación:

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

Ejemplo (Estructura de equicorrelación):

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

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

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

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

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

De este modo,

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

o equivalentemente,

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

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

Observación:

Note que podríamos 'reparametrizar' el modelo como:[5]

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

[5] ¿Cómo luce el espacio paramétrico ahora?

Ejemplo (Regresión lineal simple):

Considere el modelo de regresión:

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

Además, asumiremos que los errores satisfacen

$$\mathsf{E}(\epsilon_t) = 0, \qquad \mathsf{var}(\epsilon_t) = \sigma^2,$$

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

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

Observación:

Usualmente se asume[6] que la distribución que caracteriza los primeros momentos de $\{\epsilon_t\}$ es conocida.

[6] por ejemplo en el curso MAT266: Análisis de Regresión

Observación:

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

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

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

Ejemplo:

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

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

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

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

Modelo estadístico

Ejemplo:

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

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

de ahí que

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

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

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

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

Supuesto A0 (Identificabilidad):

Para $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ con $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. Si las distribuciones P_{θ_1} y P_{θ_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,

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

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

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

Ejemplo:

Considere, las siguientes estadísticas

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

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

Tenemos $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^2$, con $\boldsymbol{T}(X_1, \ldots, X_n) \to (T_1, T_2)$.

Ejemplo:

Considere la función de distribución empírica:

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

donde $(X_1, \ldots, X_n)^\top$ 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

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

son estadísticas para μ y σ^2, respectivamente. Mientras que

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

no es un estadístico para σ^2 pues depende de μ.