Ciencias de Datos con R: Fundamentos Estadísticos

Ana M. Bianco, Jemina García y Mariela Sued.

Método de Máxima Verosimilitud.

Inferencia Estadística

Recordemos.....

- Ingredientes: datos generados por un mecanismo aleatorio: por ej., tiramos una moneda al aires sucesivas veces.

- Objetivo: inferir *algo relacionado* con el mecanismo (aleatorio) que genera los datos, por ejemplo: ¿cuál es la probabilidad de obtener cara con nuestra moneda?

- Mecanismo: Función de distribución.
  - Caso discreto: función de probabilidad puntual
  - Caso continuo: función de densidad

- Modus Operandi: hacer *alguna cuenta* con los datos para obtener un valor que *se parezca* al que queremos estimar.

Muestra - Datos

- Muestra aleatoria: $X_1, \cdots, X_n$ variables aleatorias i.i.d. (independientes e idénticamente distribuidas)

- Datos u observaciones: $\mathbf{x} = x_1, \ldots, x_n$ constituyen una realización de la muestra aleatoria.

Inferencia Estadística

- Muestra: $(X_i)_{i \geq 1}$ i.i.d. $X_i \sim F$, $F \in \mathcal{F}$ familia de distribuciones posibles para nuestro problema
- Objetivo: inferir *algo relacionado* con el mecanismo que genera los datos:
  - $\mathbb{E}_F[X_1]$
  - $\mathbb{V}_F(X_1)$
  - $\mathbb{P}_F(X_1 \leq 40)$
  - F.
- Proponer un estimador para cada uno de los *mongo* planteados.

La empírica

Sean $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim F$. Definimos

$$\widehat{F}_n(t) := \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq t\}}$$

- $\widehat{F}_n(t)$ es una función aleatoria.
- $\widehat{F}_n(t)$ representa a una acumulada que da peso $1/n$ a $X_1, X_2, \ldots, X_n$.
- Ley de los Grandes Números:

$$\lim_{n \to \infty} \widehat{F}_n(t) = F(t) \ , \quad \text{en probabilidad}$$

Estimación Plug-in

Posibles mongos:

- $\textbf{mongo}_1(F) = \mathbb{E}_F(X)$
- $\textbf{mongo}_2(F) = \mathbb{V}_F(X)$
- $\textbf{mongo}_3(F) = \mathbb{P}_F(X \leq 40)$

El procedimiento **plug-in** propone
estimar **mongo**(F) con **mongo**$(\widehat{F}_n)$

Inferencia Estadística

- Muestra: $(X_i)_{i\geq 1}$ i.i.d. $X_i \sim F$, $F \in \mathcal{F}$ familia de distribuciones posibles para nuestro problema

- Objetivo: inferir *algo relacionado* con el mecanismo que genera los datos:
  - $\mathbb{E}_F[X_1]$
  - $\mathbb{V}_F(X_1)$
  - $\mathbb{P}_F(X_1 \leq 40)$
  - F.
- Proponer un estimador para cada uno de los *mongo* planteados.
- *Enfoque Funcional No Paramétrico:* **mongo**$(F) = T(F)$

$$T(F) \text{ lo estimás por } T(\hat{F}_n)$$

Vayamos a las tareas de Clase Parte 1.

Modelos Paramétricos: Ejemplos discretos

- Bernoulli: $X \sim \mathcal{B}(1, \theta)$
  $p(x, \theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$.
  $\theta \in \Theta = [0, 1] \subset \mathbb{R}$.

Modelos Paramétricos: Ejemplos discretos

- Bernoulli: $X \sim \mathcal{B}(1, \theta)$
  $p(x, \theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$.
  $\theta \in \Theta = [0, 1] \subset \mathbb{R}$.

- Poisson: $X \sim P(\theta)$
  $p(x, \theta) = e^{-\theta} \theta^x / x!$, $x = 0, 1, 2, \ldots$.
  $\theta \in \Theta = \mathbb{R}_{>0}$.

Modelos Paramétricos: Ejemplos discretos

- Bernoulli: $X \sim \mathcal{B}(1, \theta)$
  $p(x, \theta) = \theta^x (1 - \theta)^{1-x}$, $x = 0, 1$.
  $\theta \in \Theta = [0, 1] \subset \mathbb{R}$.

- Poisson: $X \sim P(\theta)$
  $p(x, \theta) = e^{-\theta} \theta^x / x!$, $x = 0, 1, 2, \ldots$.
  $\theta \in \Theta = \mathbb{R}_{>0}$.

$$\mathcal{M} = \{p(\cdot, \theta)\ , \theta \in \Theta\}\ ,$$

Modelos Paramétricos: Ejemplos continuos

- Uniforme: $X \sim \mathcal{U}(0, \theta)$
  $f(x, \theta) = \frac{1}{\theta}\, \mathcal{I}_{[0,\theta]}(x)$
  $\theta \in \Theta = \mathbb{R}_{>0}$.

- Exponencial: $X \sim \mathcal{E}(\theta)$
  $f(x, \theta) = \theta e^{-\theta x} \mathcal{I}_{[0,\infty)}(x)$
  $\theta \in \Theta = \mathbb{R}_{>0}$.

- Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$
  $f(x, \theta) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$
  $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$

Modelos Paramétricos: Ejemplos continuos

- Uniforme: $X \sim \mathcal{U}(0, \theta)$
  $f(x, \theta) = \frac{1}{\theta}\, \mathcal{I}_{[0,\theta]}(x)$
  $\theta \in \Theta = \mathbb{R}_{>0}$.

- Exponencial: $X \sim \mathcal{E}(\theta)$
  $f(x, \theta) = \theta e^{-\theta x} \mathcal{I}_{[0,\infty)}(x)$
  $\theta \in \Theta = \mathbb{R}_{>0}$.

- Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$
  $f(x, \theta) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$
  $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$

$$\mathcal{M} = \{f(\cdot, \theta)\ , \theta \in \Theta\}\ .$$

Modelos Paramétricos: $\mathcal{M}$ Ejemplos

Asumimos que la función de distribución F que genera los datos pertenece a una familia conocida, salvo por el valor de cierto parámetro:

$$\mathcal{M} = \{F(\cdot, \theta) \, , \theta \in \Theta\} \, ,$$

siendo $\Theta \subset \mathbb{R}^k$, para algún k.

- Caso discreto: $\mathcal{M} = \{p(\cdot, \theta) \, , \theta \in \Theta\} \, ,$
- Caso continuo: $\mathcal{M} = \{f(\cdot, \theta) \, , \theta \in \Theta\} \, .$

Si $F \in \mathcal{M}$, para conocerla necesitamos identificar su θ

Verosimilitud: Idea

- Antes de realizar el *experimento* el **resultado** es **desconocido**.
- Las probabilidades nos permiten predecir un resultado **desconocido** a partir de parámetros **conocidos**: por ejemplo

$$P(resultado|\theta) \overset{\text{por ej. caso Bernoulli}}{=} P(x|\theta) = \theta^x(1-\theta)^{(1-x)}$$

Verosimilitud: Idea

- Antes de realizar el *experimento* el **resultado** es **desconocido**.
- Las probabilidades nos permiten predecir un resultado **desconocido** a partir de parámetros **conocidos**: por ejemplo

$$P(resultado|\theta) \overset{\text{por ej. caso Bernoulli}}{=} P(x|\theta) = \theta^{x}(1-\theta)^{(1-x)}$$

Ahora se invierte el paradigma:

- Al realizar el experimento el **resultado** se hace **conocido**: **dato**.
- Nos interesa cuán **verosímil** es que un determinado parámetro haya generado el dato.

Verosimilitud: Ejemplo

- Monedas, cara$= 1$, ceca$= 0$.
- Moneda bolsillo derecho: equilibrada
- Moneda bolsillo izquierdo: probabilidad de cara es 0.8.
- Objetivo: identificar la moneda a partir de una muestra. En $n = 100$ lanzamientos se observa la muestra

$$\mathbf{x} = \underbrace{1,\ldots,1}_{\textbf{12 veces}},\underbrace{0,\ldots,0}_{\textbf{5 veces}},\underbrace{1,\ldots,1}_{\textbf{23 veces}},\underbrace{0,\ldots,0}_{\textbf{8 veces}},\underbrace{1,\ldots,1}_{\textbf{15 veces}},\underbrace{0,\ldots,0}_{\textbf{3 veces}},$$
$$\underbrace{1,\ldots,1}_{11\, veces},\underbrace{0,\ldots,0}_{4\, veces},\underbrace{1,\ldots,1}_{13\, veces},\underbrace{0,\ldots,0}_{6\, veces}$$

- ¿Cuál de las dos monedas diría que está utilizando?

Función de Verosimilitud

- $L(0.8; \mathbf{x}) =$ probabilidad de observar la muestra $\mathbf{x}$ con moneda de $p = 0.8$.

- $L(0.5; \mathbf{x}) =$ probabilidad de observar la muestra $\mathbf{x}$ con moneda de $p = 0.5$ (equilibrada).

Función de Verosimilitud

- $L(0.8;\mathbf{x})$ = probabilidad de observar la muestra $\mathbf{x}$ con moneda de $p = 0.8$.

$$
\begin{aligned}
L(0.8;\mathbf{x}) \;&=\; \underbrace{0.8\ldots0.8}_{12\,veces}\underbrace{0.2\ldots0.2}_{5\,veces}\underbrace{0.8\ldots0.8}_{23\,veces}\underbrace{0.2\ldots0.2}_{8\,veces}\underbrace{0.8\ldots0.8}_{15\,veces}\underbrace{0.2\ldots0.2}_{3\,veces}\\
&\quad \underbrace{0.8\ldots0.8}_{11\,veces}\underbrace{0.2\ldots0.2}_{4\,veces}\underbrace{0.8\ldots0.8}_{13\,veces}\underbrace{0.2\ldots0.2}_{6\,veces} =\\
&\quad = (0.8)^{74}(0.2)^{26} = 4.523128\;10^{-26}
\end{aligned}
$$

siendo 74 el número de caras observadas en las $n = 100$ repeticiones.

¿Y ahora?

- $\mathbf{x}$ con 74 caras, 26 cecas.
- $L(0.8;\mathbf{x}) = 0.8^{74}0.2^{26} = 4.523128\ 10^{-26}$
- $L(0.5;\mathbf{x}) \ = \ (1/2)^{74}(1/2)^{26} = 7.888609\ 10^{-31}$
- ¿cuál de las dos monedas diríamos que se está utilizando?

Propuesta de Máxima Verosimilitud

La propuesta de máxima verosimilitud consiste en pensar que la moneda que estamos utilizando es aquella para la cual los valores observados resultan más probables. Es decir, elegimos la moneda que maximiza la probabilidad de los valores observados. Siendo que

$$L(0.8; \mathbf{x}) \;>\; L(0.5; \mathbf{x})$$

<u>concluimos que se está utilizando la moneda no equilibrada</u>.

Sintetizando

- $X_i \sim \mathcal{B}(1, \theta)$, $\theta \in [0, 1]$.
- 74 caras en $n = 100$ repeticiones.
- ¿θ?
- vamos a elegir aquel valor del parámetro para el cuál los valores observados tienen mayor probabilidad de ocurrir.

Función de Verosimilitud

- $L(\theta; \mathbf{x})$: Mide cuál es la probabilidad de observar nuestra realización $\mathbf{x}$ cuando la probabilidad de cara es θ.

- en $\mathbf{x}$ de tamaño 100 hay 74 caras

$$L(\theta; \mathbf{x}) \ = \ \theta^{74}(1-\theta)^{26}$$

- queremos maximizar $L(\theta)$.

- $\iff$ maximizar $l(\theta; \mathbf{x}) = \ln(\mathbf{L}(\theta; \mathbf{x}))$

- $l(\theta; \mathbf{x}) = 74 \ln(\theta) + 26 \ln(1-\theta)$, se maximiza en $74/100$.

- Tenemos así que el valor estimado del parámetro con la muestra dada es $= 0.74$.