Algo màs sobre Estadística

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

Estimadores de máxima verosimilitud

Introducción

En la clase anterior, vimos que uno de los problemas centrales de la estadística es la estimación de parámetros de una distribución.

Supongamos que tenemos una población y queremos medir una cierta variable aleatoria, cuya distribución F no conocemos, pero sabemos o suponemos que $F \in \mathcal{F}$, una cierta familia de distribuciones.

Para estimar un parámetro $\theta = \theta(F)$, tomamos una muestra eleatoria de tamaño n de nuestra población. Esto nos dará variables

$$X_1, X_2, \ldots, X_n$$

todas con distribución F e independientes. Entonces queremos estimar θ mediante un estimador

$$\widehat{\theta}(X_1, X_2, \ldots, X_n)$$

En esta clase, veremos un método general para obtener estimadores con buenas propiedades.

Notación vectorial

En muchos ejemplos la distribución estará caracterizada por un número finito k de parámetros, que podemos pensar como componentes de un vector

$$\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \mathbb{R}^k$$

que se mueve en una cierta región $A \subset \mathbb{R}^k$ de parámetros admisibles.

Por ejemplo, podemos pensar en la familia de distribuciones normales:

$$\mathcal{F} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$$

En este caso $\theta = (\mu, \sigma) \in A$, donde

$$A = \{(mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$$

En general, podemos escribir

$$\mathcal{F} = \{F_\theta : \theta \in A\}$$

Caso discreto

Comenzemos considerando el caso discreto. En ete caso la distribución F_θ vendrá dadas por las probabilidades puntuales, que dependerán del vector de parámetros θ:

$$p_\theta(x) = P\{X_i = x\} \quad \text{(las mismas para todo)}$$

(que serán cero salvo para numerables valores de de x)
Por ejemplo, supongamos que tenemos una urna con un cierto número de bolitas blancas B y otro tanto de rojas R, y que extraemos n bolitas con reposición pero no conocemos cuántas bolitas de cada color hay. Definimos las variables aleatorias de Bernoulli

$$X_i = \begin{cases} 1 & \text{si} \quad \text{si sale roja} \\ 0 & \text{si} \quad \text{si sale blanca} \end{cases}$$

Entonces $X_i \sim \mathrm{Be}(\theta)$ donde $\theta = \frac{R}{B+R} \in [0,1] = A$. Aquí los posibles valores de las X_i son 0 y 1, y sus pobablidades

$$p_\theta(1) = \theta, \quad p_\theta(0) = 1 - \theta$$

Verosimilitud en el caso discreto

Ahora nos preguntamos: si el parámetro θ tuviera un cierto valor, ¿cuál sería la probabilidad de observar ciertos valores $x_1, X_2, \ldots, x_n$?. Esto vendrá dado por la función de verosimilitud

$$\begin{aligned}
\mathcal{L}(\theta) = \mathcal{L}(x_1, x_2, \ldots, x_n; \theta) &:= P_\theta\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \\
&= \prod_{i=1}^{n} P_\theta\{X = x_i\} \quad \text{por independencia} \\
&= \prod_{i=1}^{n} p_\theta(x_i)
\end{aligned}$$

Aquí usamos la notación P_θ para indicar que las probabilidades indicadas dependen del parámetro θ.

Verosimilitud en el caso continuo

Cuando trabajamos con variables continuas, la distribución F_θ estará caracterizada por una densidad de probabilidad f_θ. entonces definimos la función de función de verosimilitud como la densidad conjunta del vector aleatorio $(X_1, X_2, \ldots, X_n)$ correspondiente a un determinado valor del parámetro θ, que de nuevo por la independencia de la muestra será:

$$\mathcal{L}(\theta) = \mathcal{L}(x_1, x_2, \ldots, x_n; \theta) := \prod_{i=1}^{n} f_\theta(x_i)$$

Caso discreto (3)

Para cada muestra particular $(x_1, \ldots, x_n)$, la estimación de máxima verosimilitud de θ es el valor $\widehat{\theta}_{VM}$ que maximiza la verosimilitud. Es decir:

$$\mathcal{L}(x_1, x_2, x_n; \widehat{\theta}_{MV}) = \max_{\theta \in A} \mathcal{L}(x_1, \ldots, x_n; \theta)$$

El estimador de máxima verosimilitud, $\widehat{\theta}_{VM}(X_1, X_2, \ldots, X_n)$, es aquél que evaluado en cada muestra particular nos da la estimación de máxima verosimilitud

$$\widehat{\theta}_{MV}(x_1, x_2, \ldots, x_n)$$

Como $\mathcal{L}$ es un producto, conviene maximizar $\ell(s) = \log\ \mathcal{L}(\theta)$.

Continuación del ejemplo en el caso discreto

En el ejemplo que vimos antes de la distribución Be(θ):

$$\mathcal{L}(\theta) = \theta^s (1-\theta)^{n-s}$$

donde

$$s = x_1 + x_2 + \ldots + x_n$$

Luego:

$$\ell(\theta) = \log\ \mathcal{L}(\theta) = s \log \theta + (n-s) \log(1-\theta)$$

$$\ell'(\theta) = s \cdot \frac{1}{\theta} - (n-s) \cdot \frac{1}{1-\theta}$$

El máximo se va a alcanzar cuando $\ell'(s) = 0$, o sea:

$$\frac{s}{\theta} = \frac{n-s}{1-\theta} \Leftarrow \frac{1-\theta}{\theta} = \frac{n-s}{s} \Leftrightarrow \frac{1}{\theta} - s = \frac{n}{s} - 1 \Leftrightarrow \theta = \frac{s}{n}$$

Así que en este caso el mejor

$$\widehat{\theta}_{MV} = \overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Estimación de los parámetros de la distribución normal

Volvamos al ejemplo de la familia de las distribuciones normales. Son distribuciones continuas con la densidad:

$$f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-(x-\mu)^2/(2\sigma^2)} \quad \theta = (\mu, \sigma)$$

Entonces:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-(x_i-\mu)^2/(2\sigma^2)} = \frac{1}{\sigma^n\sqrt{2\pi}} \prod_{i=1}^{n}\ e^{-(x_i-\mu)^2/(2\sigma^2)}$$

luego

$$\ell(\theta) = \log\ \mathcal{L}(\theta) = -n\log\sigma - \frac{1}{2}\log(2\pi) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

Estimación de los parámetros de la distribución normal (2)

Como ahora tenemos dos parámetros, para encontrar el máximo vemos donde se anulan simultáneamente ambas derivadas parciales:

$$\frac{\partial \ell}{\partial \mu}(\theta) = -\sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma^2} = 0 \Rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ell}{\partial \sigma}(\theta) = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} = 0 \Rightarrow \sigma = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right)^{1/2}$$

O sea que los estimadores de máxima verosimilitud para los parámetros de la distribución normal son:

$$\widehat{\mu} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\widehat{\sigma} = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X}_n)^2\right)^{1/2}$$