MAT-206: Sesión 7, Estimación: Método de momentos

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Suponga $X_1, \ldots, X_n$ una muestra aleatoria desde el modelo estadístico

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

Deseamos entender el mecanismo o modelo verdadero (basado en $\boldsymbol{\theta}_0 \in \Theta$) que generó los datos.

IDEA:

Desarrollar procedimientos para 'seleccionar' $\boldsymbol{\theta}$ desde los datos (estimación) así como discutir sus propiedades.

Definición 1:

Una función $\boldsymbol{T}:\mathcal{X}^n \to \Gamma$ es llamado un estimador (puntual). Este es usado para estimar $\boldsymbol{\gamma} = \boldsymbol{g}(\boldsymbol{\theta})$.

El valor $\boldsymbol{T}(\boldsymbol{x})$ es llamado estimación de $\boldsymbol{g}(\boldsymbol{\theta})$ y corresponde a una realización de la variable aleatoria $\boldsymbol{T}(\boldsymbol{X})$.

Observación:

Usualmente anotamos un estimador (y una estimación) como $\widehat{\boldsymbol{\gamma}} = \boldsymbol{T}(X_1,\dots,X_n)$ y distinguimos el método usado como $\widehat{\gamma}_{\mathsf{ML}}$, $\widehat{\gamma}_{\mathsf{MM}}$ o $\widehat{\gamma}_{\mathsf{LS}}$.

Método de momentos

Importante:

Este método **no** requiere conocer la distribución subyacente X ($\sim \mathsf{P}_\theta \in \mathcal{P}$), sino que requiere **asumir** formas específicas para sus momentos.

Considere $X_1, \dots, X_n$ una m.a.(n) desde P_θ (uni-dimensional). Es decir,

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

y sea

$$\mu_k = \mu_k(\mathsf{P}_\theta) = \mathsf{E}(X^k) = \int x^k \, \mathsf{d}\,\mathsf{P}_\theta = \int x^k f_X(x;\boldsymbol{\theta})\, \mathsf{d}x.$$

Asumiremos también que el parámetro de interés γ depende de $\boldsymbol{\theta}$ a través de los momentos μ_k como:

$$\gamma = h(\mu_1(\mathsf{P}_\theta), \dots, \mu_r(\mathsf{P}_\theta)),$$

donde h es una función conocida.

Definición 2 (Estimador de momentos):

Suponga $X_1, \dots, X_n$ que sigue el modelo estadístico $\{\mathsf{P}_{\theta}^{\otimes n} : \boldsymbol{\theta} \in \Theta\}$. El estimador de momentos es definido como

$$\widehat{\gamma}_{\mathsf{MM}} = h(m_1, \dots, m_r),$$

donde $m_k = \widehat{\mu}_k$ es el momento empírico (o muestral) de orden k, dado por

$$m_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k.$$

Ejemplo:

Considere $X_1, \dots, X_n$ una muestra aleatoria desde P_θ y considere

$$\gamma = \int x f(x;\theta)\,\mathrm{d}x.$$

Usando el método de momentos, tenemos que:

$$\widehat{\gamma}_{\mathsf{MM}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Ejemplo:

Considere que el parámetro de interés es la desviación estándar σ. Note que

$$\sigma^2 = \int (x - \mu_1)^2 f(x; \boldsymbol{\theta})\,\mathrm{d}x,$$

de este modo,

$$\sigma = \sqrt{\mu_2 - \mu_1^2} = h(\mu_1, \mu_2),$$

cuyo estimador usando el método de momentos adopta la forma:

$$\widehat{\sigma}_{\mathsf{MM}} = \sqrt{m_2 - m_1^2}.$$

Note que

$$\begin{aligned} m_2 - m_1^2 &= \frac{1}{n}\sum_{i=1}^n x_i^2 - \Big(\frac{1}{n}\sum_{i=1}^n x_i\Big)^2 = \frac{1}{n}\Big(\sum_{i=1}^n x_i^2 - \frac{1}{n}(n\overline{x})^2\Big) \\ &= \frac{1}{n}\Big(\sum_{i=1}^n x_i^2 - n\overline{x}^2\Big) = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2. \end{aligned}$$

De este modo,

$$\widehat{\sigma}_{\mathsf{MM}} = \Big\{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2\Big\}^{1/2}.$$

Ejemplo:

El sesgo de una variable aleatoria X con distribución F es definida como

$$\gamma = \frac{\mathsf{E}_F(X - \mathsf{E}_F(X))^3}{(\mathsf{var}_F(X))^{3/2}} = \frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}},$$

así el estimador de momentos asume la forma

$$\widehat{\gamma}_{\mathsf{MM}} = \frac{m_3 - 3m_2m_1 + 2m_1^3}{(m_2 - m_1^2)^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^3}{\{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2\}^{3/2}}$$

Ejemplo:

La función de distribución acumulada es definida como

$$F(t) = \mathsf{P}_F((-\infty, t]),$$

para t fijo. Un estimador natural para la probabilidad del conjunto $(-\infty, t]$ es la frecuencia relativa,

$$\widehat{F}_n(t) = \widehat{F}_n(t; \boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(x_i).$$

$\widehat{F}_n$ se denomina la función de distribución empírica. Note que

$$m_k = \mu_k(\widehat{F}_n) = \int x^k \, \mathsf{d}\widehat{F}_n,$$

es decir, $\widehat{F}_n$ es un estimador de momentos de F.

Ejemplo:

Sea $X_1,\dots,X_n$ una muestra aleatoria desde una distribución log-normal con vector de parámetros $\boldsymbol{\theta} = (\mu,\sigma^2)^\top \in \mathbb{R}_+ \times \mathbb{R}_+$. Es decir, cada X_i tiene densidad

$$f(z;\mu,\sigma^2) = \frac{1}{\sigma z\sqrt{2\pi}} \exp\Big\{-\frac{(\log z-\mu)^2}{2\sigma^2}\Big\}, \quad z>0.$$

En este caso

$$\mu_1 = \exp(\mu+\sigma^2/2), \qquad \mu_2 = \exp(\sigma^2)\big(\exp(\mu+\sigma^2/2)\big)^2,$$

y portanto los estimadores de momentos son dados por:

$$\widehat{\mu}_{\textsf{MM}} = 2\log m_1 - \tfrac{1}{2}\log m_2, \qquad \widehat{\sigma}^2_{\textsf{MM}} = \log m_2 - 2\log m_1.$$

Ejemplo:

Suponga $X_1, \dots, X_n$ una muestra aleatoria desde $\mathrm{Poi}(\lambda)$, $\lambda > 0$. Recuerde que

$$\mathsf{E}(X_1) = \mathsf{var}(X_1) = \lambda.$$

De este modo, podemos considerar

$$\widehat{\lambda}_{\mathsf{MM}} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \big(= h(m_1)\big)$$

por otro lado, otro estimador puede ser[1]

$$\widetilde{\lambda}_{\mathsf{MM}} = h(m_1, m_2) = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

[1] ¿Cuál $\widehat{\lambda}_{\mathsf{MM}}$ o $\widetilde{\lambda}_{\mathsf{MM}}$ es mejor?

En la práctica, tenemos que $\boldsymbol{\theta}$ es vector p-dimensional. Por simplicidad asumiremos que $\boldsymbol{\gamma} = \boldsymbol{\theta}$. De este modo tenemos

$$\begin{aligned} \mu_1 &= g_1(\theta_1, \dots, \theta_p), \\ &\vdots \\ \mu_p &= g_p(\theta_1, \dots, \theta_p). \end{aligned}$$

Resolviendo para los p-parámetros en función de los momentos, obtenemos

$$\begin{aligned} \theta_1 &= h_1(\mu_1, \dots, \mu_p), \\ &\vdots \\ \theta_p &= h_p(\mu_1, \dots, \mu_p). \end{aligned} \tag{1}$$

El estimador $\widehat{\boldsymbol{\theta}}_{\mathsf{MM}}$, puede ser obtenido substituyendo en (1) por los momentos muestrales, es decir:

$$\begin{aligned}
\widehat{\theta}_1 &= h_1(m_1,\dots,m_p),\\
&\vdots\\
\widehat{\theta}_p &= h_p(m_1,\dots,m_p).
\end{aligned}$$

Debemos resaltar que el método de momentos requiere resolver el sistema de ecuaciones **no lineales**

$$\begin{aligned}
g_1(\theta_1,\dots,\theta_p) - \mu_1 &= 0,\\
&\vdots\\
g_p(\theta_1,\dots,\theta_p) - \mu_p &= 0,
\end{aligned}$$

El sistema de ecuaciones puede ser escrito como:

$$\boldsymbol{\Psi}(\boldsymbol{\theta}) = \mathbf{0}, \tag{2}$$

donde $\boldsymbol{\Psi} : \mathbb{R}^p \rightarrow \mathbb{R}^p$ y por tanto $\widehat{\boldsymbol{\theta}}_{\mathsf{MM}}$ corresponde a una raíz de la ecuación de estimación en (2).

En general, para obtener $\widehat{\boldsymbol{\theta}}_{\mathsf{MM}}$ se requiere de métodos iterativos, tal como:

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} - \{\dot{\boldsymbol{\Psi}}(\boldsymbol{\theta}^{(s)})\}^{-1}\boldsymbol{\Psi}(\boldsymbol{\theta}^{(s)}), \qquad s = 0, 1, \dots,$$

donde $\boldsymbol{\theta}^{(s)}$ representa una estimación para $\boldsymbol{\theta}$ en la etapa s-ésima y

$$\dot{\boldsymbol{\Psi}}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\Psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}.$$

Método de momentos

Observación:

Una dificultad evidente del método de momentos es que $\boldsymbol{\Psi}(\boldsymbol{\theta}) = \mathbf{0}$ puede tener múltiples raíces.

Importante:

El método de momentos es un caso particular de una función de inferencia. Más adelante, se estudiará las propiedades de estimadores obtenidos como la raíz de una función de inferencia.