MAT-206: Sesión 10, Propiedades asintóticas

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Definición 1 (Consistencia débil):

Una secuencia de estimadores $\{T_n\}$ para el parámetro $\gamma = g(\theta)$ es débilmente consistente, si T_n converge en probabilidad a γ, esto es, si para cualquier $\epsilon > 0$ y para todo $\theta \in \Theta$,

$$\lim_{n\to\infty} \mathsf{P}_\theta(|T_n - g(\theta)| > \epsilon) = 0,$$

y escribimos $T_n \xrightarrow{\mathsf{P}} \gamma$.

Definición 2 (Consistencia fuerte):

Se dice que $\{T_n\}$ converge con probabilidad 1 o casi seguramente (a.s.) a γ, para todo $\theta \in \Theta$, esto es,

$$\mathsf{P}_\theta\left(\lim_{n\to\infty} T_n = g(\theta)\right) = 1,$$

es decir T_n es consistente fuerte, en cuyo caso anotamos $T_n \xrightarrow{\text{a.s.}} \gamma$.

Resultado 1 (Teorema del mapeo continuo):

Sea $\{S_n\}$ una secuencia de variables aleatorias, S_0 una variable aletoria y h una función continua.

i) Si $S_n \xrightarrow{\text{P}} S_0$, entonces
$$h(S_n) \xrightarrow{\text{P}} h(S_0).$$

ii) Si $S_n \xrightarrow{\text{a.s.}} S_0$, entonces
$$h(S_n) \xrightarrow{\text{a.s.}} h(S_0).$$

Demostración:

Una prueba puede ser hallada en Gut (2005).[1]

[1] Probability: A Graduate Course. Springer, New York.

Propiedades asintóticas

Una herramienta importante para la verificación de consistencia es la desigualdad de Chebyshev. Para una variable aleatoria Z con media finita, tenemos

$$\mathsf{P}(|Z - \mathsf{E}(Z)| > \tau) \leq \frac{\mathsf{var}(Z)}{\tau^2}.$$

Ejemplo:

Sea $\{X_n\}$ una secuencia de variables aleatorias IID con función de distribución F. Entonces se tiene:

- La media aritmética converge a $\mathsf{E}_F(X) = \mu$, es decir, $\overline{X}_n \xrightarrow{\mathsf{P}} \mu$. En efecto, usando la desigualdad de Chebyshev y asumiendo $\sigma < \infty$, tenemos

  $$\mathsf{P}(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\mathsf{var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0,$$

  para $n \to \infty$.

Ejemplo:

- La varianza empírica y la desviación estándar convergen a $\text{var}_F(X) = \sigma^2$ y σ, respectivamente:
$$S_n^2 \xrightarrow{\mathsf{P}} \sigma^2, \qquad S_n \xrightarrow{\mathsf{P}} \sigma.$$

- La frecuencia relativa de un evento A converge a su probabilidad. Sea
$$Q_n(A) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in A),$$
donde $I(\cdot)$ denota la función indicadora. Entonces $Q_n(A) \xrightarrow{\mathsf{P}} \mathsf{P}(A)$.

- Si el j-ésimo momento $\mu_j = \mathsf{E}_F(X^j)$ existe, entonces los momentos empíricos
$$m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j,$$
son consistentes. Es decir, $m_j \xrightarrow{\mathsf{P}} \mu_j$. De ahí que, el estimador de momentos para $\gamma = h(\mu_1, \dots, \mu_r)$ es un estimador consistente (para h contínua).

Definición 3 (Convergencia en distribución):

Sea $\{X_n\}$ una secuencia de variables aleatorias y sea X otra variable aleatoria. Además, considere F_n la CDF de X_n y F la CDF de X. Se dice que X_n converge en distribución a X, en cuyo caso escribimos $X_n \xrightarrow{\mathsf{D}} X$ si

$$\lim_{n\to\infty} F_n(t) = F(t),$$

para todo t donde F es continua.

Resultado 2:

Sea $\{Z_n\}$ una secuencia de variables aleatorias y sea $M_n(t)$ la MGF de Z_n. Sea Z una variable aleatoria con MGF dada por $M(t)$. Si

$$M_n(t) \to M(t), \qquad \text{para todo } |t| < h,\ h > 0.$$

Entonces, $Z_n \xrightarrow{\mathsf{D}} Z$.

Resultado 3 (Teorema de Slutsky):

Considere dos secuencias de variables aleatorias $\{X_n\}$, $\{Y_n\}$, una variable aleatoria X y una constante fija c. Suponga que $X_n \xrightarrow{\mathsf{D}} X$, y $Y_n \xrightarrow{\mathsf{P}} c$. Entonces:

i) $X_n \pm Y_n \xrightarrow{\mathsf{D}} X \pm c$.

ii) $X_n Y_n \xrightarrow{\mathsf{D}} cX$.

iii) $X_n Y_n^{-1} \xrightarrow{\mathsf{D}} c^{-1} X$ siempre que $\mathsf{P}(Y_n = 0) = 0$ para todo n y $c \neq 0$.

Definición 4 (Normalidad asintótica):

Una secuencia de estimadores $\{\boldsymbol{T}_n\}$ para el parámetro m-dimensional $\boldsymbol{\gamma} = \boldsymbol{g}(\boldsymbol{\theta})$ es asintóticamente normal si para todo $\boldsymbol{\theta} \in \Theta$ la distribución de $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{g}(\boldsymbol{\theta}))$ converge a una distribución normal con media cero y matriz de covarianza $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Es decir,

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{g}(\boldsymbol{\theta})) \xrightarrow{\mathsf{D}} \mathsf{N}_m(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})).$$

Observación:

Un estimador es asintóticamente eficiente si este es asintóticamente normal, con

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \Big(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big)\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta})\Big(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big)^\top,$$

donde $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta})$ es la matriz de información de la distribución subyacente.

Resultado 4 (Método Delta):

Suponga $\boldsymbol{T}_n$ un estimador de la forma $\boldsymbol{T}_n = \boldsymbol{h}(\boldsymbol{S}_n)$ donde la secuencia $\{\boldsymbol{S}_n\}$ es asintóticamente normal, esto es,

$$\sqrt{n}(\boldsymbol{S}_n - \boldsymbol{\mu}) \xrightarrow{\mathsf{D}} \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

para $\boldsymbol{\mu} \in \mathbb{R}^k$ y $\boldsymbol{\Sigma} > 0$. Si $\partial \boldsymbol{h}(\boldsymbol{\mu})/\partial \boldsymbol{\mu}^\top$ es matriz de rango completo, entonces

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{h}(\boldsymbol{\mu})) \xrightarrow{\mathsf{D}} \mathsf{N}\Big(\boldsymbol{0}, \Big(\frac{\partial \boldsymbol{h}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^\top}\Big)\boldsymbol{\Sigma}\Big(\frac{\partial \boldsymbol{h}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^\top}\Big)^\top\Big).$$

Ejemplo:

Sean $X_1, \dots, X_n$ variables aleatorias IID con $\mathsf{E}(X_i) = \mu \neq 0$ y $\mathsf{var}(X_i) = \sigma^2 < \infty$. El parámetro $\gamma = \log \mu$ es estimado por $\widehat{\gamma}_n = \log \overline{X}_n$. Este estimador es consistente y asintóticamente normal. En efecto,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathsf{D}} \mathsf{N}(0, \sigma^2).$$

Como $h(s) = \log s$ y $h'(s) = 1/s$, sigue que

$$\sqrt{n}(\log \overline{X}_n - \log \mu) \xrightarrow{\mathsf{D}} \mathsf{N}\Big(0, \frac{\sigma^2}{\mu^2}\Big).$$

Considere una secuencia de estimadores $\{\widehat{\boldsymbol{\theta}}_n\}$ que converge en probabilidad a $\boldsymbol{\theta}_0$ perteneciendo al interior de Θ, y suponga que

Supuesto A5:

La matriz

$$\boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta}_0) = \mathsf{E}_{\theta_0}\left\{-\frac{\partial^2 \log f(\boldsymbol{y}_1;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\},$$

existe y es no singular.

Resultado 5 (Distribución asintótica del MLE):

Bajo las condiciones A1-A5 una secuencia de máximos locales de $\ell_n(\boldsymbol{\theta})$ tiene distribución asintótica

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta}_0)).$$

Distribución asintótica del MLE

Demostración:

Note que la secuencia $\{\widehat{\boldsymbol{\theta}}_n\}$ satisface las ecuaciones de verosimilitud, $\partial\ell_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$. Usando una expansión de Taylor del vector score en torno de $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, resulta[2]

$$\frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\partial\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \frac{\partial^2\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_{\mathsf{P}}(\mathbf{1}),$$

haciendo $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n$, sigue que

$$\mathbf{0} = \frac{\partial\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \frac{\partial^2\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_{\mathsf{P}}(\mathbf{1}),$$

es decir

$$-\frac{\partial^2\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{\partial\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + o_{\mathsf{P}}(\mathbf{1}).$$

[2] Si $Z_n \xrightarrow{\mathsf{P}} 0$ entonces anotamos $Z_n = o_{\mathsf{P}}(1)$.

Evidentemente,

$$\Big(-\frac{1}{n}\frac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+o_{\mathsf{P}}(\mathbf{1}).$$

Por otro lado, tenemos que

$$-\frac{1}{n}\frac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}=\frac{1}{n}\sum_{i=1}^n -\frac{\partial^2 \log f(\boldsymbol{Y}_i;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$$

converge casi seguramente a

$$\boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta}_0)=\mathsf{E}_{\theta_0}\Big(-\frac{\partial^2 \log f(\boldsymbol{Y}_i;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big),$$

debido a la ley fuerte de los grandes números. Lo anterior lleva a,

$$\boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+o_{\mathsf{P}}(\mathbf{1}),$$

Note también que,

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(\boldsymbol{Y}_i;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\partial \log f(\boldsymbol{Y}_i;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \mathsf{E}_{\theta_0}\left(\frac{\partial \log f(\boldsymbol{Y}_i;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)\right\},\end{aligned}$$

converge en distribución a

$$\mathsf{N}_k(\mathbf{0}, \mathsf{Cov}_{\theta_0}(\boldsymbol{U}_i(\boldsymbol{\theta}_0))) \stackrel{\mathsf{d}}{=} \mathsf{N}_k(\mathbf{0}, \boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta}_0)).$$

Premultiplicando por $\boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta}_0)$ sigue que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta}_0)\boldsymbol{U}_n(\boldsymbol{\theta}_0) + o_{\mathsf{P}}(\mathbf{1}).$$

Distribución asintótica del MLE

Usando que

$$\frac{1}{\sqrt{n}}\boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta}_0)\boldsymbol{U}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta}_0)),$$

sigue el resultado deseado.

Observación:

Bajo condiciones apropiadas el MLE es consistente, asintóticamente normal y eficiente (BUE). Además, este resultado permite el desarrollo de intervalos de confianza y test de hipótesis asintóticos.