

MAT-206: 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 $\theta \in \Theta$,

$$\lim_{n \rightarrow \infty} P_{\theta}(|T_n - g(\theta)| > \epsilon) = 0,$$

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

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,

$$P_{\theta} \left(\lim_{n \rightarrow \infty} T_n = g(\theta) \right) = 1,$$

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



Definición 3 (Convergencia en r -ésima media):

Se dice que una secuencia de vectores $\{\mathbf{T}_n\}$ **converge en r -ésima media** ($r > 0$) a un vector aleatorio \mathbf{T} (posiblemente degenerado) si

$$\lim_{n \rightarrow \infty} E_{\theta}(\|\mathbf{T}_n - \mathbf{T}\|^r) = 0,$$

y escribimos $\mathbf{T}_n \xrightarrow{r\text{th}} \mathbf{T}$.

Definición 4 (Consistencia en media cuadrática):

Una secuencia de estimadores $\{\mathbf{T}_n\}$ para el parámetro $\gamma = g(\theta)$ es **consistente en media cuadrática**,¹ si

$$\lim_{n \rightarrow \infty} E_{\theta}(\|\mathbf{T}_n - g(\theta)\|^2) = 0,$$

en cuyo caso escribimos $\mathbf{T}_n \xrightarrow{2\text{nd}} \gamma$.

¹Es decir, $MSE(\mathbf{T}_n, \theta) \rightarrow 0$ cuando $n \rightarrow \infty$.



Resultado 1 (Teorema del mapeo continuo):

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

i) Si $S_n \xrightarrow{P} S_0$, entonces

$$h(S_n) \xrightarrow{P} h(S_0).$$

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

$$h(S_n) \xrightarrow{a.s.} h(S_0).$$

Demostración:

Una prueba puede ser hallada en Gut (2005).²

²Probability: A Graduate Course. Springer, New York.



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

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

Ejemplo:

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

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

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

para $n \rightarrow \infty$.



Ejemplo:

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

$$S_n^2 \xrightarrow{P} \sigma^2, \quad S_n \xrightarrow{P} \sigma.$$

- ▶ 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 $Q_n(A) \xrightarrow{P} P(A)$.

- ▶ Si el j -ésimo momento $\mu_j = E_F(X^j)$ existe, entonces los momentos empíricos

$$m_j = \frac{1}{n} \sum_{i=1}^n X_i^j,$$

son consistentes. Es decir, $m_j \xrightarrow{P} \mu_j$. De ahí que, el estimador de momentos para $\gamma = h(\mu_1, \dots, \mu_r)$ es un estimador consistente (para h continua).



Definición 5 (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 $X_n \xrightarrow{D} X$ si

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

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

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

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



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 $X_n \xrightarrow{D} X$, y $Y_n \xrightarrow{P} c$. Entonces:

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



Definición 6 (Normalidad asintótica):

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

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{D} N_m(0, \Sigma(\theta)).$$

Observación:

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

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

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



Resultado 4 (Método Delta):

Suponga T_n un estimador de la forma $T_n = h(S_n)$ donde la secuencia $\{S_n\}$ es asintóticamente normal, esto es,

$$\sqrt{n}(S_n - \mu) \xrightarrow{D} N(0, \Sigma),$$

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

$$\sqrt{n}(T_n - h(\mu)) \xrightarrow{D} N\left(0, \left(\frac{\partial h(\mu)}{\partial \mu^\top}\right) \Sigma \left(\frac{\partial h(\mu)}{\partial \mu^\top}\right)^\top\right).$$



Ejemplo:

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

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

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

$$\sqrt{n}(\log \bar{X}_n - \log \mu) \xrightarrow{D} N\left(0, \frac{\sigma^2}{\mu^2}\right).$$



Considere una secuencia de estimadores $\{\hat{\theta}_n\}$ que converge en probabilidad a θ_0 perteneciendo al interior de Θ , y suponga que

Supuesto A5:

La matriz

$$\mathcal{F}_1(\theta_0) = E_{\theta_0} \left\{ - \frac{\partial^2 \log f(\mathbf{y}_1; \theta_0)}{\partial \theta \partial \theta^\top} \right\},$$

existe y es no singular.



Resultado 5 (Distribución asintótica del MLE):

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

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N_k(\mathbf{0}, \mathcal{F}_1^{-1}(\theta_0)).$$



Demostración:

Note que la secuencia $\{\hat{\theta}_n\}$ satisface las ecuaciones de verosimilitud, $\partial \ell_n(\theta) / \partial \theta = \mathbf{0}$. Usando una expansión de Taylor del vector score en torno de $\theta = \theta_0$, resulta³

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{\partial \ell_n(\theta_0)}{\partial \theta} + \frac{\partial^2 \ell_n(\theta_0)}{\partial \theta \partial \theta^\top} (\theta - \theta_0) + o_P(1),$$

haciendo $\theta = \hat{\theta}_n$, sigue que

$$\mathbf{0} = \frac{\partial \ell_n(\theta_0)}{\partial \theta} + \frac{\partial^2 \ell_n(\theta_0)}{\partial \theta \partial \theta^\top} (\hat{\theta}_n - \theta_0) + o_P(1),$$

es decir

$$-\frac{\partial^2 \ell_n(\theta_0)}{\partial \theta \partial \theta^\top} (\hat{\theta}_n - \theta_0) = \frac{\partial \ell_n(\theta_0)}{\partial \theta} + o_P(1).$$

³Si $Z_n \xrightarrow{P} 0$ entonces anotamos $Z_n = o_P(1)$.

Distribución asintótica del MLE

Evidentemente,

$$\left(-\frac{1}{n} \frac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_P(1).$$

Por otro lado, tenemos que

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

converge casi seguramente a

$$\mathcal{F}_1(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0} \left(-\frac{\partial^2 \log f(\mathbf{Y}_i; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right),$$

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

$$\mathcal{F}_1(\boldsymbol{\theta}_0) \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_P(1),$$



Note también que,

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

converge en distribución a

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

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

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



Usando que

$$\frac{1}{\sqrt{n}} \mathcal{F}_1^{-1}(\theta_0) U_n(\theta_0) \xrightarrow{D} N_k(\mathbf{0}, \mathcal{F}_1^{-1}(\theta_0) \mathcal{F}_1(\theta_0) \mathcal{F}_1^{-1}(\theta_0)),$$

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.

