Consistencia de los estimadores de momentos y de máxima verosimilitud
Distribución asintótica de los estimadores de máxima verosimilitud

Graciela Boente[1]

[1]Universidad de Buenos Aires and CONICET, Argentina

Sea $\mathcal{P} = \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ una familia de distribuciones.

Supongamos que para cada n se tiene un estimador $\delta_n(X_1, \ldots, X_n)$ de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n.

Sea $\mathcal{P} = \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ una familia de distribuciones.

Supongamos que para cada n se tiene un estimador $\delta_n(X_1, \ldots, X_n)$ de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n.

Definición. $\delta_n(X_1, \ldots, X_n)$ es una *sucesión fuertemente consistente de estimadores* de $q(\boldsymbol{\theta})$ si

$$\lim_{n\to\infty} \delta_n(X_1, \ldots, X_n) = q(\boldsymbol{\theta}) \qquad c.t.p.$$

o sea si $P_{\boldsymbol{\theta}}(\delta_n(X_1, \ldots, X_n) \to q(\boldsymbol{\theta})) = 1$ para todo $\boldsymbol{\theta} \in \Theta$.

Sea $\mathcal{P} = \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ una familia de distribuciones.

Supongamos que para cada n se tiene un estimador $\delta_n(X_1, \ldots, X_n)$ de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n.

Definición. $\delta_n(X_1, \ldots, X_n)$ es una *sucesión fuertemente consistente de estimadores* de $q(\boldsymbol{\theta})$ si

$$\lim_{n\to\infty} \delta_n(X_1, \ldots, X_n) = q(\boldsymbol{\theta}) \qquad c.t.p.$$

o sea si $P_{\boldsymbol{\theta}}(\delta_n(X_1, \ldots, X_n) \to q(\boldsymbol{\theta})) = 1$ para todo $\boldsymbol{\theta} \in \Theta$.

Definición. $\delta_n(X_1, \ldots, X_n)$ es una *sucesión débilmente consistente de estimadores* de $q(\boldsymbol{\theta})$ si

$$\delta_n(X_1, \ldots, X_n) \xrightarrow{p} q(\boldsymbol{\theta})$$

o sea, si para todo $\varepsilon > 0$ y $\boldsymbol{\theta} \in \Theta$

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}}(|\delta_n(X_1, \ldots, X_n) - q(\boldsymbol{\theta})| > \varepsilon) = 0\,.$$

Propiedades

Proposición 1. Sea, para todo n, $\delta_n = \delta_n(X_1, \ldots, X_n)$ un estimador de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n. Si

- $\mathrm{VAR}_{\boldsymbol{\theta}}(\delta_n) \to 0$
- $\mathbb{E}_{\boldsymbol{\theta}}(\delta_n) \to q(\boldsymbol{\theta})$,

entonces $\delta_n(X_1, \ldots, X_n)$ es débilmente consistente para $q(\boldsymbol{\theta})$.

Propiedades

Proposición 1. Sea, para todo n, $\delta_n = \delta_n(X_1, \ldots, X_n)$ un estimador de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n. Si

- $\mathrm{VAR}_{\boldsymbol{\theta}}(\delta_n) \to 0$
- $\mathbb{E}_{\boldsymbol{\theta}}(\delta_n) \to q(\boldsymbol{\theta})$,

entonces $\delta_n(X_1, \ldots, X_n)$ es débilmente consistente para $q(\boldsymbol{\theta})$.

Proposición 2. Sea $\delta_n(X_1, \ldots, X_n)$ una sucesión de estimadores IMVU para $q(\boldsymbol{\theta})$, donde $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución perteneciente a la familia $\mathcal{P} = \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$. Luego,

- $\mathrm{VAR}_{\boldsymbol{\theta}}(\delta_n(X_1, \ldots, X_n)) \to 0$ para todo $\boldsymbol{\theta} \in \Theta$

de donde $\delta_n(X_1, \ldots, X_n)$ es débilmente consistente para $q(\boldsymbol{\theta})$.

Consistencia de los estimadores de los momentos

Teorema 1.

- Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución perteneciente a la familia $\mathcal{P} = \{P_\theta,\ \theta \in \Theta \subset \mathbb{R}\}$,
- $g(x)$ una función con valores en $\mathbb{R}$
- Supongamos que $M(\theta) = \mathbb{E}_\theta(g(X_1))$ es, como función de θ, continua y estrictamente monótona.

Sea el estimador de momentos $\widehat{\theta}_n$ definido como la solución de

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = M(\theta) = \mathbb{E}_\theta(g(X_1)). \tag{1}$$

Luego con probabilidad 1 existe n_0 tal que para todo $n \geq n_0$ la ecuación (1) tiene una solución, $\widehat{\theta}_n$, y $\widehat{\theta}_n$ es fuertemente consistente para θ.

Consistencia de los estimadores de los máxima verosimilitud

Teorema 2. Sea $X_1, \ldots, X_n$ i.i.d. $X_i \sim f(x, \theta)$ con $\theta \in \Theta$, donde Θ es un intervalo abierto de $\mathbb{R}$.
Supongamos que:

- $f(x, \theta)$ es derivable respecto de θ
- el conjunto $\mathcal{S} = \{x : f(x, \theta) \neq 0\}$ es independiente de θ.
- $\theta_1 \neq \theta_2$ implica que $f(x, \theta_1) \neq f(x, \theta_2)$.

Sea $\widehat{\theta}_n$ el estimador de máxima verosimilitud de θ, que satisface

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i, \widehat{\theta}_n)}{\partial \theta} = 0 \tag{2}$$

Supongamos además que la ecuación (2) tiene a lo sumo una solución. Entonces,

$$\widehat{\theta}_n \xrightarrow{a.s.} \theta$$

o sea, $\widehat{\theta}_n$ es una sucesión de estimadores fuertemente consistente.

Lema previo

Sean p y q dos densidades o dos funciones de densidad discretas o continuas distintas. Luego se tiene:

$$\mathbb{E}_p \left[\log \left(\frac{q(X)}{p(X)} \right) \right] < 0$$

donde $\mathbb{E}_p$ significa que se calcula la esperanza considerando que X tiene una distribución discreta o continua cuya densidad o probabilidad puntual es p.

Estimadores asintóticamente normales

Sea $\mathcal{P} = \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ una familia de distribuciones.

Supongamos que para cada n se tiene un estimador $\delta_n(X_1, \ldots, X_n)$ de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n.

Estimadores asintóticamente normales

Sea $\mathcal{P} = \{P_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ una familia de distribuciones.

Supongamos que para cada n se tiene un estimador $\delta_n(X_1, \ldots, X_n)$ de $q(\boldsymbol{\theta})$ basado en una muestra aleatoria de tamaño n.

Definición 1. Diremos que $\delta_n(X_1, \ldots, X_n)$ es una sucesión de *estimadores asintóticamente normal* si

$$\sqrt{n}(\delta_n(X_1, \ldots, X_n) - q(\boldsymbol{\theta})) \xrightarrow{D} N\left(0, \sigma^2(\boldsymbol{\theta})\right) \tag{3}$$

Estimadores asintóticamente normales y eficientes

Definición 2. Supongamos que

- $X_i \sim f(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$
- El conjunto $\mathcal{S} = \{x : f(x, \theta) > 0\}$ es independiente de θ.
- Para todo x, $f(x, \theta)$ es derivable respecto de θ.

$$0 < \mathrm{I}_1(\theta) = \mathbb{E}_\theta \left(\frac{\partial \log f(X, \theta)}{\partial \theta} \right)^2 < \infty$$

Se dice que $\delta_n(X_1, \ldots, X_n)$ es una sucesión de *estimadores asintóticamente normal y eficiente* (A.N.E.) o simplemente una sucesión de *estimadores asintóticamente eficientes* si

- $\delta_n(X_1, \ldots, X_n)$ es una sucesión de *estimadores asintóticamente normal*, o sea, cumple (3) y
- su varianza asintótica es igual a

$$\sigma^2(\theta) = \frac{[q'(\theta)]^2}{\mathrm{I}_1(\theta)}$$

Propiedad

Sea $\{Z_n\}_{n\geq 1}$, una sucesión de variables aleatorias tales que

$$\sqrt{n}\,(Z_n - \lambda) \xrightarrow{D} N(0, \sigma^2)\ .$$

Sea $H : \mathbb{R} \to \mathbb{R}$ una función de clase $\mathcal{C}^1$ tal que $H'(\lambda) \neq 0$.

Entonces

$$\sqrt{n}\,(H(Z_n) - H(\lambda)) \xrightarrow{D} N\left(0, \sigma^2 \left[H'(\lambda)\right]^2\right)\ .$$

Dist. Asint. de los estimadores de los momentos

Teorema 3.

- Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución perteneciente a la familia $\mathcal{P} = \{P_\theta,\ \theta \in \Theta \subset \mathbb{R}\}$,
- Supongamos que $M(\theta) = \mathbb{E}_\theta(g(X_1))$ es, tal que $M : \Theta \to \Lambda$ es estrictamente monótona.

  Sea $H : \Lambda \to \Theta$ su inversa, $H(t) = M^{-1}(t)$.
- Supongamos que H es continuamente diferenciable y tal que $H'(\mu) \neq 0$ donde $\mu = M(\theta)$

Sea el estimador de momentos $\widehat{\theta}_n$ definido como la solución de (1)

Entonces,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{D} N\left(0, \left(H'(\mu)\right)^2\ \mathrm{VAR}_\theta g(X_1)\right)$$

Teorema 4.

Supongamos que $\mathbf{X}_1, \ldots \mathbf{X}_n$ son i.i.d. $\mathbf{X}_i \sim f(\mathbf{x}, \theta)$, con

- $\theta \in \Theta \subset \mathbb{R}$,
- Θ abierto

y que se cumple

(A) El conjunto $\mathcal{S} = \{\mathbf{x} : f(\mathbf{x}, \theta) > 0\}$ es independiente de θ.

(B) Para todo $\mathbf{x}$, $f(\mathbf{x}, \theta)$ es tres veces derivable respecto de θ y

$$\frac{\partial^3 f(\mathbf{x}, \theta)}{\partial \theta^3}$$

es continua en θ.

(C) Si $h(\mathbf{X})$ es un estadístico tal que $\mathbb{E}_\theta |h(\mathbf{X})| < \infty$ para todo $\theta \in \Theta$ entonces $g(\theta) = \int h(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x}$ es dos veces diferenciable y para $s = 1, 2$

$$\frac{\partial^s}{\partial \theta^s} \int h(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x} = \int h(\mathbf{x}) \frac{\partial^s f(\mathbf{x}, \theta)}{\partial \theta^s} d\mathbf{x}$$

Teorema 4.

(D) $0 < \mathrm{I}_1(\theta) = \mathbb{E}_\theta \left(\dfrac{\partial \log f(\mathbf{X}, \theta)}{\partial \theta} \right)^2 < \infty$

(E) Para cada θ_0 fijo, existen $c > 0$ y una función $M(\mathbf{x}) > 0$ tales que

$$\begin{aligned} \left| \frac{\partial^3 \log f(\mathbf{x}, \theta)}{\partial \theta^3} \right| = \left| \frac{\partial^2 \psi(\mathbf{x}, \theta)}{\partial \theta^2} \right| & \leq M(\mathbf{x}) \\ \mathbb{E}_{\theta_0} M(\mathbf{X}_1) & < \infty \end{aligned}$$

para todo $\mathbf{x} \in \mathcal{S}$ y para todo $\theta \in \{\theta \in \Theta : |\theta_0 - \theta| < c\}$

Teorema 4.

Teorema 4. Bajo (A), (B), (C), (D) y (E). Sean

- $\widehat{\theta}_n$ un estimador de máxima verosimilitud de θ consistente y
- $q(\theta)$ derivable con $q'(\theta) \neq 0$ para todo θ.

Entonces, $q(\widehat{\theta}_n)$ es A.N.E. para estimar $q(\theta)$, o sea,

$$\sqrt{n}\left(q(\widehat{\theta}_n) - q(\theta)\right) \xrightarrow{D} N\left(0, \frac{[q'(\theta)]^2}{\mathrm{I}_1(\theta)}\right)$$

Teorema (Le Cam, 1953)

Supongamos que $\mathbf{X}_1, \ldots \mathbf{X}_n$ son i.i.d. $\mathbf{X}_i \sim f(\mathbf{x}, \theta)$, con

- $\theta \in \Theta \subset \mathbb{R}$,
- Θ abierto

y que se cumple

(A) El conjunto $\mathcal{S} = \{\mathbf{x} : f(\mathbf{x}, \theta) > 0\}$ es independiente de θ.

(B) Para todo $\mathbf{x}$, $f(\mathbf{x}, \theta)$ es dos veces derivable respecto de θ y

$$\frac{\partial^2 f(\mathbf{x}, \theta)}{\partial \theta^2}$$

es continua en θ.

(C) Si $h(\mathbf{X})$ es un estadístico tal que $\mathbb{E}_\theta |h(\mathbf{X})| < \infty$ para todo $\theta \in \Theta$ entonces $g(\theta) = \int h(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x}$ es dos veces diferenciable y para $s = 1, 2$

$$\frac{\partial^s}{\partial \theta^s} \int h(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x} = \int h(\mathbf{x}) \frac{\partial^s f(\mathbf{x}, \theta)}{\partial \theta^s} d\mathbf{x}$$

Teorema (Le Cam, 1953)

(D) $0 < \mathrm{I}_1(\theta) = \mathbb{E}_\theta \left(\dfrac{\partial \log f(\mathbf{X}, \theta)}{\partial \theta} \right)^2 < \infty$

(E) Para cada θ_0 fijo, existen $c > 0$ y una función $M(\mathbf{x}) > 0$ tales que

$$\begin{aligned} \left| \frac{\partial^3 \log f(\mathbf{x}, \theta)}{\partial \theta^3} \right| = \left| \frac{\partial^2 \psi(\mathbf{x}, \theta)}{\partial \theta^2} \right| &\leq M(\mathbf{x}) \\ \mathbb{E}_{\theta_0} M(\mathbf{X}_1) &< \infty \end{aligned}$$

para todo $\mathbf{x} \in \mathcal{S}$ y para todo $\theta \in \{\theta \in \Theta : |\theta_0 - \theta| < c\}$

Teorema (Le Cam)

Sea $\delta_n(X_1, \ldots, X_n)$ una sucesión de estimadores asintóticamente normales de $q(\theta)$, con q derivable y $q'(\theta) \neq 0$ para todo θ. Es decir,

$$\sqrt{n}\left(\delta_n(X_1, \ldots, X_n) - q(\theta)\right) \xrightarrow{D} N\left(0, \sigma^2(\theta)\right)$$

entonces

$$\sigma^2(\theta) \geq \frac{[q'(\theta)]^2}{\mathrm{I}_1(\theta)}$$

excepto en un conjunto $\Theta_0 \subset \Theta$ tal que Θ_0 tiene medida 0.

Ejemplo

Sea $X_1, \ldots, X_n$ i.i.d. $X_i \sim N(\theta, 1)$. Luego, $\mathrm{I}_1(\theta) = 1$. Definamos

$$\delta_n(X_1, \ldots, X_n) = \begin{cases} \overline{X} & \text{si} \quad |\overline{X}| \geq n^{-\frac{1}{4}} \\ a\,\overline{X} & \text{si} \quad |\overline{X}| < n^{-\frac{1}{4}} \end{cases}$$

Entonces

$$\sqrt{n}\,(\delta_n(X_1, \ldots, X_n) - \theta) \xrightarrow{D} N(0, \sigma^2(\theta))$$

donde

$$\sigma^2(\theta) = \begin{cases} 1 & \text{si} \quad \theta \neq 0 \\ a^2 & \text{si} \quad \theta = 0 \end{cases}$$

Por lo tanto, si $a < 1$ la desigualdad se viola en $\theta = 0$.