Estimadores:
Comportamiento Asintótico de un Estimador

Repasemos....

$\mathbf{X}_n = (X_1 \dots X_n)$ m.a. $X_i \sim F$, $F \in \mathcal{F}$, $\mathcal{F} = \{F(\cdot, \theta), \theta \in \Theta\}$.
Queremos estimar $q(\theta)$.

Sea $T_n = T(\mathbf{X}_n)$ un estimador de $q(\theta)$

- $\mathbb{E}_\theta(T_n) \Rightarrow$ sesgo $\mathbb{B}_\theta(T_n) = \mathbb{E}_\theta(T_n) - q(\theta)$

- $ECM_\theta(T_n) = \mathbb{V}ar_\theta(T_n) + \mathbb{B}^2_\theta(T_n)$

- Error standard: $se(T_n) = \sqrt{\mathbb{V}ar_\theta(T_n)}$

Recordemos: Estimación de una proporción

Queremos estimar a p, la proporción de habitantes de Argentina que está a favor de las vacunas. Para ello, elegiremos n habitantes al azar y se les preguntará si están a favor de las vacunas o no. ¿Cómo estimamos a p?

$$X_i = \begin{cases} 1 \text{ si el } i-\text{ésimo encuestado dice que sí} \\ 0 \text{ si no} \end{cases}$$

- $X_1, \ldots, X_n$ son variables aleatorias i.i.d. con distribuición $B(1, p)$, siendo p la verdadera proporción que queremos estimar.

- Un estimador puntual (EMM o EMV) de p es $\hat{p}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

- ¿Por qué tomamos $n = 1000$? ¿Sería más conveniente tomar $n = 2000$? Si es mejor una elección que otra, ¿donde se ve?

- $se(\hat{p}_n) = \sqrt{\mathbb{V}ar_p(\hat{p}_n)} = \sqrt{\dfrac{p(1-p)}{n}}$

Distribución de muestreo y error standard

- La distribución de T_n se llama **distribución de muestreo** (sample distribution).

- La desviación estandar de T_n se llama **error standard**, es por lo general desconocida y se utiliza un estimador del error estandar, lo notamos $\widehat{se}(T_n)$.

Consistencia

$\mathcal{F} = \{F(\cdot, \theta)\ , \theta \in \Theta\}$ una familia de distribuciones.
$\mathbf{X}_n = (X_1 \ldots X_n)$ m.a. $X_i \sim F$, $F \in \mathcal{F}$
$T_n = T_n(\mathbf{X}_n)$ estimador de $q(\theta)$ basado en la muestra aleatoria de tamaño n $\mathbf{X}_n$.

Def. T_n es una sucesión fuertemente consistente de estimadores de $q(\theta)$ si

$$\lim_{n\to\infty} T_n = q(\theta) \qquad c.t.p.$$

o sea si $P_\theta\left(w \in \Omega : \lim_{n\to\infty} T_n = q(\theta)\right) = 1$ para todo $\theta \in \Theta$.

Notamos: $T_n \xrightarrow{c.s.} q(\theta)$

Consistencia

$\mathcal{F} = \{F(\cdot, \theta)\ , \theta \in \Theta\}$ una familia de distribuciones.
$\mathbf{X}_n = (X_1 \ldots X_n)$ m.a. $X_i \sim F$, $F \in \mathcal{F}$
$T_n = T_n(\mathbf{X}_n)$ estimador de $q(\theta)$ basado en la muestra aleatoria de tamaño n $\mathbf{X}_n$.

Def. T_n es una sucesión débilmente consistente de estimadores de $q(\theta)$ si

$$T_n \xrightarrow{P} q(\theta)$$

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

$$\lim_{n\to\infty} P_\theta(|T_n - q(\theta)| > \varepsilon) = 0\,.$$

¿Quién puede ayudarnos? Ley de los Grandes Números

- **Ley fuerte:** Sean $\{X_n\}_{n\geq 1}$ variables aleatorias independientes tales que $E(X_j^2) < \infty$. Sea $(b_n)_{n\geq 1}$ una sucesión numérica tales que $b_n > 0$, b_n es creciente y $\lim\limits_{n\to\infty} b_n = +\infty$.

  Entonces

$$\sum_{n=1}^{\infty} \frac{V(X_n)}{b_n^2} < \infty \Longrightarrow \frac{1}{b_n} \sum_{j=1}^{n} \left(X_j - E\left(X_j\right)\right) \xrightarrow{c.s.} 0 \, .$$

- **Otra versión**: Sean $X_1, X_2, \ldots$ v.a. independientes de a pares e idénticamente distribuidas con $E\left|X_1\right| < \infty$. Si $E(X_1) = \mu$, entonces cuando $n \to \infty$

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{c.s.} \mu \, .$$

¿Quién puede ayudarnos? Ley de los Grandes Números

- **Ley débil:** Sean $\{X_n\}_{n\geq 1}$ variables aleatorias no correlacionadas, es decir $\mathbb{C}ov(X_i, X_j) = 0$ para $i \neq j$, tales que $\mathbb{E}(X_j) = \mu_j$ y $\mathbb{V}ar(X_j) = \sigma_j^2$.

  Denotemos $\overline{\mu}_n = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} \mu_j = \mathbb{E}(\overline{X}_n)$.

  Entonces, si $\lim_{n\to\infty} \dfrac{1}{n^2}\displaystyle\sum_{j=1}^{n} \sigma_j^2 = 0$, tenemos que

$$\overline{X}_n - \overline{\mu}_n \xrightarrow{p} 0 \; .$$

Aplicación en Estadística

Dada una muestra aleatoria $X_1, \ldots, X_n$ independientes, tenemos que:

- Primer momento: por la L.G.N. fuerte si $E\,|X_i| < \infty$, tenemos en forma directa un estimador consistente para $\mu = E(X_i)(= \mu_1)$, ya que si tomamos $\hat{\mu} = \bar{X}_n$, tenemos que

$$\bar{X}_n \xrightarrow{c.s.} \mu$$

y por lo tanto,

$$\bar{X}_n \xrightarrow{P} \mu$$

- Segundo momento: también da un estimador consistente para, por ejemplo, $\theta = E(X_i^2)(= \mu_2) < \infty$, ya que como antes

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E(X_i^2) = \theta$$

Recordemos algunas propiedades útiles

Algunas transformaciones preservan la convergencia casi segura y en probabilidad.

Si g es una función continua en μ, entonces

- $X_n \xrightarrow{c.s.} \mu \Rightarrow g(X_n) \xrightarrow{c.s.} g(\mu)$

Además, si $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$,

- $X_n \pm Y_n \xrightarrow{c.s.} X \pm Y$
- $X_n Y_n \xrightarrow{c.s.} XY$
- si además $\mathbb{P}(Y = 0) = 0$, $X_n / Y_n \xrightarrow{c.s.} X/Y$

Estos resultados también son válidos en probabilidad.

Estimación de $\sigma^2 = \mathbb{V}ar_F(X_i)$

- ¿Cómo estimamos σ^2?

- Un estimador posible es E.M.M.:

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

Estimación de σ^2

- Por la L.G.N., $\bar{X}_n \xrightarrow{c.s.} \mu$ y $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{c.s.} \mathbb{E}(X_1^2)$

- Como las funciones continuas preservan la convergencia c.s., $\bar{X}_n \xrightarrow{c.s.} \mathbb{E}(X_1)$, luego $\bar{X}_n^2 \xrightarrow{c.s.} \mathbb{E}^2(X_1)$ y en consecuencia

Estimación de σ^2

- Por la L.G.N., $\bar{X}_n \xrightarrow{c.s.} \mu$ y $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{c.s.} \mathbb{E}(X_1^2)$

- Como las funciones continuas preservan la convergencia c.s., $\bar{X}_n \xrightarrow{c.s.} \mathbb{E}(X_1)$, luego $\bar{X}_n^2 \xrightarrow{c.s.} \mathbb{E}^2(X_1)$ y en consecuencia

$$\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 \xrightarrow{c.s.} \mathbb{E}(X_1^2) - \mathbb{E}^2(X_1) = \mathbb{V}ar(X_1) = \sigma^2$$

Estimación de σ^2

- Por la L.G.N., $\bar{X}_n \xrightarrow{c.s.} \mu$ y $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{c.s.} \mathbb{E}(X_1^2)$

- Como las funciones continuas preservan la convergencia c.s., $\bar{X}_n \xrightarrow{c.s.} \mathbb{E}(X_1)$, luego $\bar{X}_n^2 \xrightarrow{c.s.} \mathbb{E}^2(X_1)$ y en consecuencia

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \xrightarrow{c.s.} \mathbb{E}(X_1^2) - \mathbb{E}^2(X_1) = \mathbb{V}ar(X_1) = \sigma^2$$

- Varianza Muestral: $S_n^2 = \frac{n}{n-1}\widehat{\sigma}_n^2$.

  Como la sucesión $\frac{n}{n-1} \to 1$, S_n^2 también resulta consistente.

Error cuadrático medio y consistencia

Teorema Sea T_n un estimador de $q(\theta)$ basado en una muestra aleatoria de tamaño n. Si $ECM_\theta(T_n) \to 0$ cuando $n \to \infty$, entonces T_n es un estimador débilmente consistente de $q(\theta)$.

Demo:
Recordemos la **desigualdad de Markov:** Sean $h : \mathbb{R} \to \mathbb{R}^+$ tal que h es par y restringida a $\mathbb{R}^+$ es creciente y además Y una v.a. tal que $\mathbb{E}(h(Y))$ existe. Entonces, para todo $\epsilon > 0$

$$\mathbb{P}(|Y| \geq \epsilon) \leq \frac{\mathbb{E}[h(Y)]}{h(\epsilon)}$$

Error cuadrático medio y consistencia

Teorema Sea T_n un estimador de $q(\theta)$ basado en una muestra aleatoria de tamaño n. Si $ECM_\theta(T_n) \to 0$ cuando $n \to \infty$, entonces T_n es un estimador débilmente consistente de $q(\theta)$.

Demo:
Recordemos la **desigualdad de Markov:** Sean $h : \mathbb{R} \to \mathbb{R}^+$ tal que h es par y restringida a $\mathbb{R}^+$ es creciente y además Y una v.a. tal que $\mathbb{E}(h(Y))$ existe. Entonces, para todo $\epsilon > 0$

$$\mathbb{P}(|Y| \geq \epsilon) \leq \frac{\mathbb{E}[h(Y)]}{h(\epsilon)}$$

Dado $\epsilon > 0$, por la desigualdad de Markov tenemos
$$0 \leq \mathbb{P}\left(|T_n - q(\theta)| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}_\theta\left[\left(|T_n - q(\theta)|\right)^2\right] = \frac{1}{\varepsilon^2} ECM_\theta(T_n)$$

Corolario

Corolario. Sea $T_n = T_n(X_1, \ldots, X_n)$ un estimador de $q(\theta)$ basado en una muestra aleatoria de tamaño n. Si cuando $n \to \infty$

- $\mathbb{E}_\theta(T_n) \to q(\theta)$,
- $\mathbb{V}ar_\theta(T_n) \to 0$

entonces T_n es débilmente consistente para $q(\theta)$.

Ejemplo: Modelo de posición con errores normales

Un investigador desea determinar cuanto vale cierta magnitud y para ello realiza n mediciones.
Supone que

$$X_i = \theta + \epsilon_i, \ i = 1, \ldots, n$$

donde $\epsilon_i \sim N(0, \sigma_o^2)$ independientes, con σ_o^2 conocida.

¿Cómo estimamos a θ?

¿Cómo es la distribución de muestreo del estimador propuesto?

¿Y si σ_o^2 es desconocida?

Ejemplo: ¿y si los errores no son normales?

Supone que

$$X_i = \theta + \epsilon_i, \ i = 1, \ldots, n$$

donde $\mathbb{E}(\epsilon_i) = 0$ y $\mathbb{V}ar(\epsilon_i) = \sigma^2$ e independientes.

- $\hat{\theta}_n = \bar{X}_n$
- $\mathbb{V}ar(\hat{\theta}_n) = \sigma^2/n \Rightarrow se(\hat{\theta}_n) = \sqrt{\sigma^2/n}$
- $\frac{\bar{X}_n - \theta}{\sqrt{\sigma^2/n}} \sim ??$

Distribución Asintótica

Sea $T_n = T(\mathbf{X}_n) = T(X_1, \ldots, X_n)$ un estimador de θ basado en la m.a. $X_1, \ldots, X_n$.

Supongamos que para cierta sucesión numérica c_n tal que $c_n \to \infty$ y cierta variable aleatoria W, tenemos que cuando $n \to \infty$

$$c_n(T_n - \theta) \xrightarrow{\mathcal{D}} W\,,$$

entonces diremos que T_n converge en distribución a W a tasa c_n.

Distribución Asintótica

Sea $T_n = T(\mathbf{X}_n) = T(X_1, \ldots, X_n)$ un estimador de θ basado en la m.a. $X_1, \ldots, X_n$.

Supongamos que para cierta sucesión numérica c_n tal que $c_n \to \infty$ y cierta variable aleatoria W, tenemos que cuando $n \to \infty$

$$c_n(T_n - \theta) \xrightarrow{\mathcal{D}} W \,,$$

entonces diremos que T_n converge en distribución a W a tasa c_n.
Típicamente, tenemos que $c_n = \sqrt{n}$, ¿por qué?

Ejemplo: ¿y si los errores no son normales?

Supone que

$$X_i = \theta + \epsilon_i,\ i = 1, \ldots, n$$

donde $\mathbb{E}(\epsilon_i) = 0$ y $\mathbb{V}ar(\epsilon_i) = \sigma^2$ e independientes.

- $\hat{\theta}_n = \bar{X}_n$
- $\mathbb{V}ar(\hat{\theta}_n) = \sigma^2/n \Rightarrow se(\hat{\theta}_n) = \sqrt{\sigma^2/n}$
- por el T.C.L. tenemos que $\frac{\bar{X}_n - \theta}{\sqrt{\sigma^2/n}} \xrightarrow{\mathcal{D}} N(0, 1)$
- ¿Qué pasa si estimamos el $se(\hat{\theta}_n)$? Por ejemplo, ¿si usamos $\widehat{se}(\hat{\theta}_n) = \sqrt{S_n^2/n}$?

$$\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \xrightarrow{\mathcal{D}} ?$$

¿Quíen puede ayudarnos? Teorema de Slutsky

Sean $\{X_n\}_{n\geq 1}$ e $\{Y_n\}_{n\geq 1}$ sucesiones de variables aleatorias, tales que $X_n \xrightarrow{\mathcal{D}} X$ e $Y_n \xrightarrow{P} c$, entonces

- $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$
- $X_n - Y_n \xrightarrow{\mathcal{D}} X - c$
- $X_n \cdot Y_n \xrightarrow{\mathcal{D}} X \cdot c$
- Si $c \neq 0$, entonces $X_n/Y_n \xrightarrow{\mathcal{D}} X/c$

Entonces....

Supongamos que $\widehat{se}(\hat{\theta}_n)$ es un estimador consistente de $se(\hat{\theta}_n)$. Luego,

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \frac{\hat{\theta}_n - \theta}{\widehat{se}(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1)$$

Entonces....

Supongamos que $\widehat{se}(\hat{\theta}_n)$ es un estimador consistente de $se(\hat{\theta}_n)$. Luego,

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \frac{\hat{\theta}_n - \theta}{\widehat{se}(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1)$$

Demo:

$$\frac{\hat{\theta}_n - \theta}{\hat{s}e(\hat{\theta}_n)} = \frac{se(\hat{\theta}_n)}{\widehat{se}(\hat{\theta}_n)} \, \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)}$$

Entonces....

Supongamos que $\widehat{se}(\hat{\theta}_n)$ es un estimador consistente de $se(\hat{\theta}_n)$.
Luego,

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \frac{\hat{\theta}_n - \theta}{\widehat{se}(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1)$$

Demo:

$$\frac{\hat{\theta}_n - \theta}{\hat{se}(\hat{\theta}_n)} = \frac{se(\hat{\theta}_n)}{\widehat{se}(\hat{\theta}_n)} \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)}$$

Como

$$\frac{se(\hat{\theta}_n)}{\widehat{se}(\hat{\theta}_n)} \xrightarrow{P} 1 \text{ y } \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} N(0,1)$$

Entonces....

Supongamos que $\widehat{se}(\hat{\theta}_n)$ es un estimador consistente de $se(\hat{\theta}_n)$.
Luego,

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \frac{\hat{\theta}_n - \theta}{\widehat{se}(\hat{\theta}_n)} \xrightarrow{\mathcal{D}} N(0,1)$$

Demo:

$$\frac{\hat{\theta}_n - \theta}{\hat{se}(\hat{\theta}_n)} = \frac{se(\hat{\theta}_n)}{\widehat{se}(\hat{\theta}_n)} \; \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)}$$

Como

$$\frac{se(\hat{\theta}_n)}{\widehat{se}(\hat{\theta}_n)} \xrightarrow{P} 1 \text{ y } \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} N(0,1)$$

por Slutzky, se llega a lo que queríamos probar:

$$\frac{\hat{\theta}_n - \theta}{\widehat{se}(\hat{\theta}_n)} = \frac{se(\hat{\theta}_n)}{\widehat{se}(\hat{\theta}_n)} \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} N(0,1)$$

Consistencia de los Estimadores de Momentos

- Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución perteneciente a la familia $\mathcal{F} = \{F_\theta,\ \theta \in \Theta \subset \mathbb{R}\}$,

- $g(x)$ una función continua 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

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = M(\widehat{\theta}_n) \qquad (1)$$

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 Máxima Verosimilitud

Sea $X_1, \ldots, X_n$ i.i.d. $X_i \sim f(x, \theta_0)$ con $\theta_0 \in \Theta$, donde Θ es un intervalo abierto de $\mathbb{R}$. Supongamos que:

R0. $f(x, \theta)$ es derivable respecto de θ

R1. el conjunto $\mathcal{S} = \{x : f(x, \theta) \neq 0\}$ es independiente de θ.

R2. $\theta_1 \neq \theta_2$ implica que $f(x, \theta_1) \neq f(x, \theta_2)$.

Sea $\widehat{\theta}_n$ el estimador de máxima verosimilitud de θ_0, que es la única solución de

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

Entonces,

$$\widehat{\theta}_n \xrightarrow{c.s.} \theta_0.$$

Veamos un esbozo de la demostración.

Resultado previo (Video 1)

Lema: Sean p y q dos densidades o dos funciones de probabilidad distintas con soporte común. Entonces,

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

Resultado previo (Video 1)

Lema: Sean p y q dos densidades o dos funciones de probabilidad distintas con soporte común. Entonces,

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

Caso continuo. Es claro que no puede ocurrir que $q(X)/p(X) = k$ c.t.p., si fuera así, entonces $\mathbb{E}_p(q(X)/p(X)) = k$ y

$$\int_{-\infty}^{+\infty} \frac{q(x)}{p(x)} p(x)dx = k \Rightarrow \int_{-\infty}^{+\infty} q(x)dx = k$$

luego, $k = 1$ pues $q(x)$ es una densidad, contradice la hipótesis pues $p \neq q$.
Además, $h(x) = -\ln(x)$ es una función estrictamente convexa ya que $\frac{d^2(-\ln x)}{dx^2} = \frac{1}{x^2} > 0$. Aplicando la desigualdad de Jensen,

$$\mathbb{E}_p\left[-\ln \frac{q(X)}{p(X)}\right] > -\ln\left[\mathbb{E}_p \frac{q(X)}{p(X)}\right] = -\ln \int_{-\infty}^{+\infty} \frac{q(x)}{p(x)} p(x)dx = -\ln 1 = 0.$$

Bosquejo de la demo de la consistencia del EMV (Video 1)

- Log–verosimilitud: $\ell_n(\mathbf{X}_n, \theta) = \ell_n(X_1, \ldots, X_n, \theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i, \theta)$
- $\widehat{\theta}_n$ satisface $\ell_n(\mathbf{X}_n, \widehat{\theta}_n) = \max_{\theta\in\Theta} \ell_n(\mathbf{X}_n, \theta)$ y $\frac{\partial \ell_n(\mathbf{X}_n, \widehat{\theta}_n)}{\partial\theta} = 0$.
- Además, se tiene

$$\ell_n(\mathbf{X}_n, \theta_0 + \delta) - \ell_n(\mathbf{X}_n, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{f(X_i, \theta_0 + \delta)}{f(X_i, \theta_0)}\right)$$

$$\ell_n(\mathbf{X}_n, \theta_0 - \delta) - \ell_n(\mathbf{X}_n, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{f(X_i, \theta_0 - \delta)}{f(X_i, \theta_0)}\right)$$

- Aplicando el Lema resulta que

$$\mathbb{E}_{\theta_0}\left(\log\left[\frac{f(X_1, \theta_0 + \delta)}{f(X_1, \theta_0)}\right]\right) < 0 \text{ y } \mathbb{E}_{\theta_0}\left(\log\left[\frac{f(X_1, \theta_0 - \delta)}{f(X_1, \theta_0)}\right]\right) < 0$$

Por la Ley Fuerte, con probabilidad 1 existe un n_0 tal que si $n > n_0$

$$\ell_n(\mathbf{X}_n, \theta_0 + \delta) < \ell_n(\mathbf{X}_n, \theta_0)$$

$$\ell_n(\mathbf{X}_n, \theta_0 - \delta) < \ell_n(\mathbf{X}_n, \theta_0)$$

Bosquejo de la demo de la consistencia del EMV (Video 1)

Luego, para $n > n_0$ en $(\theta_0 - \delta, \theta_0 + \delta)$ existe un máximo relativo, digamos θ_n^*, que satisface

$$\frac{\partial \ell_n(\mathbf{X}_n, \theta_n^*)}{\partial \theta} = 0 \ ,$$

Como hemos supuesto que $\widehat{\theta}_n$ era el único que satisfacía esta igualdad, resulta $\widehat{\theta}_n = \theta_n^*$ y por lo tanto

$$\widehat{\theta}_n \in (\theta_0 - \delta, \theta_0 + \delta)$$