Cota de Rao–Cramer y Eficiencia

Detalles: ver Apunte de Boente y Yohai.

- Estableceremos la conocida desigualdad de Rao–Cramer que da una cota inferior para la varianza de todo estimador insesgado bajo condiciones de regularidad.

- Veremos luego que, bajo condiciones de regularidad, los EMV alcanzan esta cota asintóticamente.

- Sea $X_1, \ldots, X_n$ i.i.d. $X_i \sim f(x, \theta)$ con $\theta \in \Theta$, $X \sim X_i$, donde Θ es un intervalo abierto de $\mathbb{R}$. Supongamos además de R1 a R2 que:

R3. $f(x, \theta)$ es dos veces diferenciable respecto de θ

R4. La integral $\int f(x, \theta) dx$ es dos veces diferenciable respecto de θ y estas derivadas son intercambiables con el signo integral.

Estimador de máxima verosimilitud

Recordemos que el estimador de máxima verosimilitud de θ es el valor que maximiza

$$\ell_n(\theta, \mathbf{X}) = \sum_{i=1}^{n} \log f(X_i, \theta)$$

Estimador de máxima verosimilitud

Recordemos que el estimador de máxima verosimilitud de θ es el valor que maximiza

$$\ell_n(\theta, \mathbf{X}) = \sum_{i=1}^{n} \log f(X_i, \theta)$$

Bajo condiciones de regularidad que vimos el EMV se puede hallar como el cero de

$$\frac{\partial}{\partial\theta}\ell_n(\theta, \mathbf{X}_n) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta)$$

Estimador de máxima verosimilitud

Recordemos que el estimador de máxima verosimilitud de θ es el valor que maximiza

$$\ell_n(\theta, \mathbf{X}) = \sum_{i=1}^{n} \log f(X_i, \theta)$$

Bajo condiciones de regularidad que vimos el EMV se puede hallar como el cero de

$$\frac{\partial}{\partial \theta} \ell_n(\theta, \mathbf{X}_n) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i, \theta)$$

Función de score:

$$\frac{\partial}{\partial \theta} \log f(X, \theta)$$

Número de Información de Fisher

Notemos que $\frac{\partial}{\partial\theta}\log f(x,\theta) = \frac{1}{f(x,\theta)}\frac{\partial}{\partial\theta}f(x,\theta)$. Luego,

$$1 = \int_{-\infty}^{\infty} f(x,\theta)dx \Rightarrow 0 = \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x,\theta)dx$$

Número de Información de Fisher

Notemos que $\frac{\partial}{\partial\theta}\log f(x,\theta) = \frac{1}{f(x,\theta)}\frac{\partial}{\partial\theta}f(x,\theta)$. Luego,

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} f(x,\theta)dx \Rightarrow 0 = \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x,\theta)dx \\
0 &= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)dx \\
0 &= \int_{-\infty}^{\infty} \frac{\partial f(x,\theta)/\partial\theta}{f(x,\theta)} f(x,\theta)dx \\
0 &= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\log f(x,\theta) f(x,\theta)dx \qquad (A)
\end{aligned}$$

Luego,

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right] = 0$$

$\frac{\partial}{\partial\theta}\log f(X,\theta)$ se llama **función de score**.

Número de Información de Fisher

Derivamos ahora en ambos miembros de (A)

$$0 = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} \log f(x,\theta) f(x,\theta) dx + \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log f(x,\theta) \frac{\partial}{\partial \theta} \log f(x,\theta) f(x,\theta) dx$$

observemos que

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log f(x,\theta) \frac{\partial}{\partial \theta} \log f(x,\theta) f(x,\theta) dx = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f(X,\theta)\right)^2\right]$$

Número de Información de Fisher

Derivamos ahora en ambos miembros de (A)

$$0 = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} \log f(x,\theta) f(x,\theta) dx + \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \log f(x,\theta) \frac{\partial}{\partial\theta} \log f(x,\theta) f(x,\theta) dx$$

observemos que

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \log f(x,\theta) \frac{\partial}{\partial\theta} \log f(x,\theta) f(x,\theta) dx = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} \log f(X,\theta)\right)^2\right] = I(\theta)$$

$I(\theta)$ **número de información de Fisher de una variable aleatoria**.

Número de Información de Fisher

Derivamos ahora en ambos miembros de (A)

$$0 = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} \log f(x,\theta) f(x,\theta) dx + \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log f(x,\theta) \frac{\partial}{\partial \theta} \log f(x,\theta) f(x,\theta) dx$$

observemos que

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log f(x,\theta) \frac{\partial}{\partial \theta} \log f(x,\theta) f(x,\theta) dx = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f(X,\theta)\right)^2\right] = I(\theta)$$

$I(\theta)$ **número de información de Fisher de una variable aleatoria**.

Además tenemos que

$$I(\theta) = -\int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} \log f(x,\theta) f(x,\theta) dx = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(x,\theta)\right]$$

Ejemplo: $X \sim B(1, \theta)$

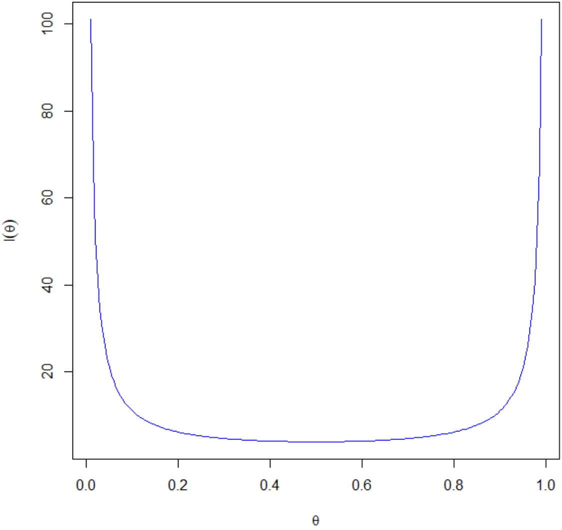
$$
\begin{aligned}
\log f(x,\theta) &= x\log\theta + (1-x)\log(1-\theta) \\
\frac{\partial}{\partial\theta}\log f(x,\theta) &= \frac{x}{\theta} - \frac{1-x}{1-\theta} \\
\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}
\end{aligned}
$$

Claramente,

$$
\begin{aligned}
I(\theta) &= -E\left[\frac{-X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right] \\
&= \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{(1-\theta)} = \frac{1}{\theta(1-\theta)}
\end{aligned}
$$

que aumenta para valores de θ cercanos a 0 y 1.

Ejemplo: $X \sim B(1, \theta)$



Algunas observaciones

- El número de información indica la curvatura de la función que maximizamos al calcular el EMV en θ.
- A mayor curvartura mayor precisión en la estimación.
- El número de información de una muestra de tamaño n, $I_n(\theta)$, es n veces el de una muestra de tamaño 1. (ver Apunte de Boente y Yohai)
- Interpretemos $I_n(\theta)$ como la cantidad de información que nos da la muestra para estimar θ.

Cota de Rao–Cramer

Bajo las condiciones de regularidad R1 a R4 , si $\widehat{\theta}_n = \widehat{\theta}_n(\mathbf{X}_n)$ es un estimador insesgado de θ, entonces

$$\mathbb{V}ar(\widehat{\theta}_n) \geq \frac{1}{nI\left(\theta\right)}$$

Cota de Rao–Cramer

Bajo las condiciones de regularidad R1 a R4 , si $\widehat{\theta}_n = \widehat{\theta}_n(\mathbf{X}_n)$ es un estimador insesgado de θ, entonces

$$\mathbb{V}ar(\widehat{\theta}_n) \geq \frac{1}{nI\left(\theta\right)}$$

Decimos que un estimador insesgado de θ, $\widehat{\theta}_n$, es eficiente sii su varianza alcanza la cota de Rao–Cramer.

Cota de Rao–Cramer

Bajo las condiciones de regularidad R1 a R4 , si $\widehat{\theta}_n = \widehat{\theta}_n(\mathbf{X}_n)$ es un estimador insesgado de θ, entonces

$$\mathbb{V}ar(\widehat{\theta}_n) \geq \frac{1}{nI\left(\theta\right)}$$

Decimos que un estimador insesgado de θ, $\widehat{\theta}_n$, es eficiente sii su varianza alcanza la cota de Rao–Cramer.

Volviendo al caso binomial:

¿Qué pasa con $\bar{X}_n$?

Distribución Asintótica del EMV

Consideremos además la siguiente condición

R5. $f(x, \theta)$ es tres veces diferenciable respecto de θ. Además, para todo $\theta \in \Theta$, existen una constante c y una función $M(x)$ tal que $\left| \frac{\partial^3}{\partial \theta^3} \log f(x, \theta) \right| \leq M(x)$, para todo $\theta_0 - c < \theta < \theta_0 + c$ y todo x en el soporte de la X y con $\mathbb{E}_{\theta_0}[M(X)] < \infty$.

Teorema (Video 2)

Si

- se cumplen R1 a R5
- $\widehat{\theta}_n$ un estimador de máxima verosimilitud consistente de θ_0,

entonces, $\widehat{\theta}_n$ es asintóticamente normal para estimar θ_0, o sea,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\mathrm{I}(\theta_0)}\right)$$

Bosquejo de la demo (Video 2)

Notamos: $\ell'(\theta) = \frac{\partial \ell_n(\theta, \mathbf{X}_n)}{\partial \theta}$ y $\ell''(\theta) = \frac{\partial^2 \ell_n(\theta, \mathbf{X}_n)}{\partial \theta^2}$

$$0 = \ell'(\widehat{\theta}_n) \approx \ell'\left(\theta_0\right) + \left(\widehat{\theta}_n - \theta_0\right) \ell''\left(\theta_0\right)$$

$$\left(\widehat{\theta}_n - \theta_0\right) \approx \frac{-\ell'\left(\theta_0\right)}{\ell''\left(\theta_0\right)}$$

$$n^{1/2}\left(\widehat{\theta}_n - \theta_0\right) \approx \frac{-n^{-1/2}\ell'\left(\theta_0\right)}{n^{-1}\ell''\left(\theta_0\right)}$$

Bosquejo de la demo (Video 2)

Primero, consideremos el numerador:

$$\left[n^{-1/2}\ell'\left(\theta_0\right)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i,\theta_0)}{\partial \theta}$$

Los sumandos son i.i.d. con

$$\mathbb{E}_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial \theta}\right] = 0$$

y

$$\mathbb{V}ar_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial \theta}\right] = \mathbb{E}_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial \theta}\right]^2 = I(\theta_0)$$

Bosquejo de la demo (Video 2)

Los sumandos son i.i.d. con

$$\mathbb{E}_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial\theta}\right] = 0$$

y

$$\mathbb{V}ar_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial\theta}\right] = I(\theta_0)$$

Luego, por el TCL,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i,\theta_0)}{\partial\theta} \xrightarrow{\mathcal{D}} N(0, I(\theta_0))$$

Bosquejo de la demo (Video 2)

Los sumandos son i.i.d. con

$$\mathbb{E}_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial \theta}\right] = 0$$

y

$$\mathbb{V}ar_{\theta_0}\left[\frac{\partial \log f(X_i,\theta_0)}{\partial \theta}\right] = I(\theta_0)$$

Luego, por el TCL,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i,\theta_0)}{\partial \theta} \xrightarrow{\mathcal{D}} N(0, I(\theta_0))$$

Consideremos el denominador. Por la L.G.N

$$-n^{-1}\ell''(\theta_0) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(X_i,\theta_0)}{\partial \theta^2} \xrightarrow{P} I(\theta_0)$$

y por Slutzky resulta que

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{I(\theta_0)}\right)$$

Distribución Asintótica del EMV: Método Delta

Sea $\{Z_n\}_{n\geq 1}$, una sucesión de variables aleatorias tales que para una sucesión numérica $a_n \uparrow \infty$

$$a_n\left(Z_n-\theta\right) \xrightarrow{\mathcal{D}} Z \ .$$

Sea $g:\mathbb{R}\to\mathbb{R}$ una función de clase $\mathcal{C}^1$ tal que $g'(\theta)\neq 0$, entonces

$$a_n\left(g(Z_n)-g(\theta)\right) \xrightarrow{\mathcal{D}} g'(\theta)Z \ .$$

Dem: Esto resulta de un desarrollo de Taylor. En efecto,

$$g\left(Z_n\right)=g(\theta)+g'(\theta_n^*)\left(Z_n-\theta\right) \ ,$$

donde θ_n^* es un punto intermedio entre Z_n y θ.

Veamos en primera instancia que entonces $\theta_n^* \xrightarrow{P} \theta$. En efecto,

$$|\theta_n-\theta_n^*| < |Z_n-\theta_n| = a_n\left|Z_n-\theta_n\right|\frac{1}{a_n} \xrightarrow{P} 0$$

y por lo tanto, por la continuidad de g' $|g'(\theta_n^*)-g'(\theta)| \xrightarrow{P} 0$. Luego, usando el Teorema de Slutsky tenemos

$$a_n\left(g\left(Z_n\right)-g(\theta)\right)=g'(\theta)a_n\left(Z_n-\theta\right) \xrightarrow{\mathcal{D}} g'(\theta)Z$$

Distribución Asintótica del EMV: Método Delta

Por ejemplo, si

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

y $g : \mathbb{R} \to \mathbb{R}$ una función de clase $\mathcal{C}^1$ tal que $g'(\theta) \neq 0$.
Entonces

$$\sqrt{n}\,(g(Z_n) - g(\theta)) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2 \left[g'(\theta)\right]^2\right)$$

Ejemplo: Volvamos al ejemplo que vimos:
Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad de la forma

$$f(x, \theta) = (1-\theta)\mathbb{I}_{\left(-\frac{1}{2},0\right)}(x) + (1+\theta)\mathbb{I}_{\left(0,\frac{1}{2}\right)}(x), \qquad -1 < \theta < 1$$

¿Cuál es la distribución asintótica del estimador de momentos generalizado que hallamos usando $m(\theta) = \mathbb{P}_\theta(X > 0)$?

Teorema

Bajo condiciones de regularidad si

- $\widehat{\theta}_n$ un estimador de máxima verosimilitud consistente de θ_0
- $q(\theta)$ derivable con $q'(\theta) \neq 0$ para todo θ,

entonces, $q(\widehat{\theta}_n)$ es asintóticamente normal para estimar $q(\theta)$, o sea,

$$\sqrt{n}\left(q(\widehat{\theta}_n) - q(\theta_0)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{[q'(\theta_0)]^2}{\mathrm{I}(\theta_0)}\right)$$

Teorema

Bajo condiciones de regularidad si

- $\widehat{\theta}_n$ un estimador de máxima verosimilitud consistente de θ_0
- $q(\theta)$ derivable con $q'(\theta) \neq 0$ para todo θ,

entonces, $q(\widehat{\theta}_n)$ es asintóticamente normal para estimar $q(\theta)$, o sea,

$$\sqrt{n}\left(q(\widehat{\theta}_n) - q(\theta_0)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{[q'(\theta_0)]^2}{\mathrm{I}(\theta_0)}\right)$$

- Decimos que una sucesión de estimadores T_n de $q(\theta)$ es asintóticamente normal y eficiente (A.N.E.) si

$$\sqrt{n}\left(T_n - q(\theta_0)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{[q'(\theta_0)]^2}{\mathrm{I}(\theta_0)}\right)$$