# MAT-206: Extremum estimator

**Felipe Osorio**

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

**Definición 1**

$\widehat{\boldsymbol{\theta}}$ es llamado un extremum estimator si existe una función objetivo $Q_n(\boldsymbol{\theta})$ tal que:

$$Q_n(\widehat{\boldsymbol{\theta}}) = \max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}), \qquad \Theta \subset \mathbb{R}^p. \tag{1}$$

*Observación:*

- La notación recalca que $Q_n(\boldsymbol{\theta})$ depende de los datos muestrales $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$.
- Una prueba de la existencia así como de la consistencia de $\widehat{\boldsymbol{\theta}}_n$ puede ser hallada en Gourieroux y Monfort (1995).[1]

[1] Statistics and Econometrics Models, Volume 2. Cambridge University Press.

*Ejemplo (Mínimos cuadrados):*

Considere el siguiente modelo

$$Y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \epsilon_i, \qquad i = 1, \dots, n,$$

donde $f$ es función conocida, $\boldsymbol{x}_i$ denota un vector de regresores, $\boldsymbol{\beta}$ es vector $p \times 1$ de parámetros desconocidos, y $\mathsf{E}(\epsilon_i) = 0$, $\mathsf{var}(\epsilon_i) = \sigma^2$, $\mathsf{Cov}(\epsilon_i, \epsilon_j) = 0$ para $i \neq j$.

Podemos escribir,

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{\beta}) + \boldsymbol{\epsilon},$$

con $\mathsf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ y $\mathsf{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$. De este modo, $\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}$ es obtenido como la solución de

$$\min_{\beta} \|\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\beta})\|^2,$$

es decir,

$$\min_{\beta} \sum_{i=1}^{n} (Y_i - f_i(\boldsymbol{\beta}))^2.$$

*Ejemplo (Función de producción):*

Considere la función de producción con elasticidad de sustitución constante (CES), dada por

$$Y_t = A_0\,[\delta_0 K_t^{-\rho_0} + (1-\delta_0) L_t^{-\rho_0}]^{-1/\rho_0} + \epsilon_t,$$

donde $Y_t$ representa la producción en el periodo $t$, $K_t$ denota el stock de capital, $L_t$ es el trabajo y $\epsilon_t$ es un disturbio aleatorio con media cero.

En este caso, $\boldsymbol{x}_t = (K_t, L_t)^\top$, $\boldsymbol{\beta} = (A_0, \delta_0, \rho_0)^\top$ y

$$f(\boldsymbol{x}_t; \boldsymbol{\beta}) = A\,[\delta K_t^{-\rho} + (1-\delta) L_t^{-\rho_0}]^{-1/\rho}.$$

Además las funciones de producción suelen añadir las siguientes condiciones $A > 0$, $0 < \delta < 1$ y $-1 < \rho$, lo que determina $\Theta$.

*Ejemplo (M-estimación):*

Un extremum estimator es un M-estimador[2] si la función objetivo es un promedio muestral del tipo:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} m(Y_i;\boldsymbol{\theta}),$$

donde $m$ es una función real valorada de $(Y_i, \boldsymbol{\theta})$.

Este procedimiento es frecuente para modelos de posición-escala, definidos por la densidad

$$\frac{1}{\sigma}\, f\Big(\frac{y - \boldsymbol{x}^\top\boldsymbol{\beta}}{\sigma}\Big),$$

en cuyo caso la función $m(\cdot)$ es escogida con el objetivo de reducir la influencia de observaciones extremas.

[2] Casos particulares: máxima verosimilitud (ML) y mínimos cuadrados no lineales (NLS)

*Ejemplo (Estimación máximo $Lq$-verosímil):*

El estimador máximo $Lq$-verosímil (ML$q$E) de $\boldsymbol{\theta}$ es definido como aquél valor $\widehat{\boldsymbol{\theta}}_q$ que maximiza la función (Ferrari y Yang, 2010):[3]

$$L_q(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_q(f(\boldsymbol{y}_i;\boldsymbol{\theta})), \qquad q > 0,$$

donde $l_q(u)$ denota el logaritmo deformado de orden $q$, cuya definición es dada por:

$$l_q(u) = \begin{cases} (u^{1-q} - 1)/(1-q), & q \neq 1, \\ \log(u), & q = 1, \end{cases}$$

con $f(\boldsymbol{y}_i;\boldsymbol{\theta})$ la función de densidad del modelo asumido para los datos.

*Observación:*

Es fácil notar que, cuando $q = 1$ se recupera el método de estimación ML.

[3] Annals of Statistics **38**, 753-783.

*Ejemplo (Modelo mal especificado):*

Asuma que disponemos de $X_1, \dots, X_n$ variables IID generadas desde $g(x;\boldsymbol{\theta})$. Aunque los datos son modelados considerando

$$\mathcal{P}_f = \{f(x;\boldsymbol{\theta})^{\otimes n} : \boldsymbol{\theta} \in \Theta\},$$

con funciones de log-verosimilitud y score asociadas

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i;\boldsymbol{\theta}), \qquad \boldsymbol{U}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \log f(x_i;\boldsymbol{\theta}),$$

Sea[4]

$$\boldsymbol{V}(\boldsymbol{\theta}) = \mathsf{E}_g\{\boldsymbol{U}_n(\boldsymbol{\theta})\boldsymbol{U}_n^\top(\boldsymbol{\theta})\}, \qquad \boldsymbol{S}(\boldsymbol{\theta}) = \mathsf{E}_g\left\{-\frac{\partial \boldsymbol{U}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right\}.$$

[4] Note que, cuando $f = g$, tenemos $\boldsymbol{V}(\boldsymbol{\theta}) = \boldsymbol{S}(\boldsymbol{\theta}) = \boldsymbol{\mathcal{F}}(\boldsymbol{\theta})$.

### Lema 1 (Existencia)

Suponga que

(i) El espacio paramétrico $\Theta$ es un subconjunto compacto de $\mathbb{R}^p$.

(ii) $Q_n(\boldsymbol{\theta})$ es continua en $\boldsymbol{\theta}$ para todo $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$.

(iii) $Q_n(\boldsymbol{\theta})$ es una función medible de los datos para todo $\boldsymbol{\theta} \in \Theta$.

Entonces existe una función medible $\widehat{\boldsymbol{\theta}}_n$ que es solución del problema

$$\max_{\theta \in \Theta} Q_n(\boldsymbol{\theta}), \qquad \Theta \subset \mathbb{R}^p.$$

*Demostración:*

ver Gourieroux y Monfort (1995), Sección 24.1.

Considere la transformación $\boldsymbol{\gamma} = \boldsymbol{\tau}(\boldsymbol{\theta})$ definida sobre $\Theta$ y sea

$$\Gamma \equiv \boldsymbol{\tau}(\Theta) \equiv \{\boldsymbol{\gamma} = \boldsymbol{\tau}(\boldsymbol{\theta}) \text{ para algún } \boldsymbol{\theta} \in \Theta\}.$$

$\boldsymbol{\tau} : \Theta \to \Gamma$ es llamada una reparametrización si es 1-a-1.

### Propiedad 1

Sea $\widetilde{Q}_n(\boldsymbol{\gamma})$ la función objetivo asociada con el modelo reparametrizado. Un extremum estimator es invariante si y solo si

$$\widetilde{Q}_n(\boldsymbol{\gamma}) = Q_n(\boldsymbol{\tau}^{-1}(\boldsymbol{\gamma})),$$

para todo $\boldsymbol{\gamma} \in \Gamma$.

*Observación:*

Existen estimadores que no son invariantes. (p.ej. GMM)

### Propiedad 2 (Consistencia)

Suponga que

(i) El parámetro verdadero $\boldsymbol{\theta}_0$ es un elemento del interior del espacio paramétrico $\Theta$.

(ii) $Q_n(\boldsymbol{\theta})$ es cóncava sobre el espacio paramétrico (para cualquier $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$).

(iii) $Q_n(\boldsymbol{\theta})$ es una función medible para todo $\boldsymbol{\theta} \in \Theta$.

Sea $\widehat{\boldsymbol{\theta}}_n$ el extremum estimator definido por (1). Si existe una función $Q_0(\boldsymbol{\theta})$ tal que

(a) $Q_0(\boldsymbol{\theta})$ es únicamente maximizada sobre $\Theta$ en $\boldsymbol{\theta}_0 \in \Theta$.

(b) $Q_n(\boldsymbol{\theta})$ converge en probabilidad a $Q_0(\boldsymbol{\theta})$ para todo $\boldsymbol{\theta} \in \Theta$.

Entonces, $\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathsf{P}} \boldsymbol{\theta}_0$.

### Propiedad 3

Sea $\{\widehat{\boldsymbol{\theta}}_n\}$ una secuencia de extremum estimators de $\boldsymbol{\theta}$ obtenidos mediante maximizar la función objetivo $Q_n(\boldsymbol{\theta})$. Suponga que $\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathsf{P}} \boldsymbol{\theta}_0$, con $\boldsymbol{\theta}_0$ en el interior del espacio paramétrico $\Theta$. Además, asuma que $Q_n(\cdot)$ es diferenciable con relación a $\boldsymbol{\theta}$.

Entonces el estimador $\widehat{\boldsymbol{\theta}}_n$ satisface (asintóticamente) las condiciones de primer orden

$$\frac{\partial Q_n(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

## Supuestos

**B1:** $Q_n(\boldsymbol{\theta})$ es dos veces diferenciable en $\boldsymbol{\theta}$.

**B2:** Sea $\boldsymbol{A}_n(\boldsymbol{\theta}) = \partial^2 Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ y asuma que

$$\boldsymbol{A}_n(\boldsymbol{\theta}) \xrightarrow{\text{a.s}} \boldsymbol{A},$$

uniformemente.

**B3:** La matriz $\boldsymbol{A}$ es no singular.

**B4:** La secuencia $\{\sqrt{n}(\partial Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta})\}$ converge en distribución a una normal estándar con vector de medias cero y matriz de covarianza $\boldsymbol{B}$. Es decir,

$$\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{\text{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{B}).$$

### Resultado 1 (Normalidad asintótica)

Bajo los supuestos de las Propiedades 2 y 3, y los Supuestos B1 a B4, tenemos

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p\big(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\big).$$

*Esbozo de la demostración:*

Considere una función objetivo aditiva:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n Q_i(\boldsymbol{\theta}),$$

de este modo,

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^n \frac{\partial Q_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \qquad \frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2 Q_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}.$$

Asuma que $\widehat{\boldsymbol{\theta}}_n$ satisface las condiciones de primer orden, es decir $\partial Q_n(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta} = \mathbf{0}$. Por el Teorema del valor medio tenemos la siguiente expansión,

$$\begin{aligned}\frac{\partial Q_n(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} &= \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \frac{\partial^2 Q_n(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\partial Q_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \left[\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 Q_i(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),\end{aligned}$$

donde $\overline{\boldsymbol{\theta}}$ es un valor medio entre $\widehat{\boldsymbol{\theta}}_n$ y $\boldsymbol{\theta}_0$. Usando la condición de primer orden

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^n \frac{\partial Q_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \left[\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 Q_i(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Asumiendo que $\frac{1}{n}\sum_{i=1}^n \partial^2 Q_i(\overline{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ es no singular, podemos escribir

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\left[\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 Q_i(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\partial Q_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}.$$

Como $\overline{\boldsymbol{\theta}}$ está entre $\widehat{\boldsymbol{\theta}}_n$ y $\boldsymbol{\theta}_0$, y siempre que $\widehat{\boldsymbol{\theta}}_n$ sea un estimador consistente, entonces $\overline{\boldsymbol{\theta}} \xrightarrow{\mathsf{P}} \boldsymbol{\theta}_0$. Además

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 Q_i(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \xrightarrow{\text{a.s.}} \mathsf{E}\left[\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right] = \boldsymbol{A}.$$

Por otro lado,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial Q_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{B}).$$

Finalmente por el Teorema de Slutzky, sigue que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p(0, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}).$$