

MAT-466: Funciones de inferencia

Felipe Osorio

`fosorios.mat.utfsm.cl`

Departamento de Matemática, UTFSM



Configuración:

Vamos a suponer que el vector aleatorio \mathbf{Y} puede ser escrito como

$$\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top,$$

donde los \mathbf{Y}_i 's son independientes. Asumiremos también que \mathbf{Y} sigue un modelo estadístico

$$\mathcal{P} = \{P_\beta : \beta \in \mathcal{B}\}, \quad \mathcal{B} \in \mathbb{R}^k,$$

y \mathcal{Y} es el espacio muestral.

Definición 1:

Una función $\Psi_n : \mathcal{B} \times \mathcal{Y} \rightarrow \mathbb{R}^k$ tal que $\Psi_n(\beta; \cdot)$ es medible para todo $\beta \in \mathcal{B}$ se dice una **función de estimación**.



Observación:

Para una función de inferencia Ψ_n y una muestra $\mathbf{Y} \in \mathcal{Y}$ dadas, es posible obtener un estimador $\hat{\beta} = \hat{\beta}(\mathbf{Y})$ como [solución de la ecuación](#)

$$\Psi_n(\beta; \mathbf{Y}) = \mathbf{0}.$$

Ejemplo (Regresión lineal):

Sea Y_1, \dots, Y_n variables independientes tal que $Y_i \sim N(\mathbf{x}_i^\top \beta, \sigma^2)$ donde \mathbf{x}_i es vector de covariables asociada al i -ésimo individuo y β coeficientes de regresión. El estimador de mínimos cuadrados es

$$\Psi_n(\beta; \mathbf{Y}) = \sum_{i=1}^n \mathbf{x}_i (Y_i - \mathbf{x}_i^\top \beta).$$



Ejemplo (Regresión no-lineal):

Considere Y_1, \dots, Y_n independientes con $Y_i \sim N(f(\mathbf{x}_i; \boldsymbol{\beta}), \sigma^2)$ donde $f(\cdot; \boldsymbol{\beta})$ es una función suave de $\boldsymbol{\beta}$. La función de estimación es dada por

$$\boldsymbol{\Psi}_n(\boldsymbol{\beta}; \mathbf{Y}) = \sum_{i=1}^n \left(\frac{\partial f(\mathbf{x}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) (Y_i - f(\mathbf{x}_i; \boldsymbol{\beta})).$$

La función $\boldsymbol{\Psi}_n(\boldsymbol{\beta}; \mathbf{Y})$ es fácil de estudiar debido a que es lineal en los **residuos**

$$r_i(\boldsymbol{\beta}) = Y_i - f(\mathbf{x}_i; \boldsymbol{\beta}), \quad i = 1, \dots, n.$$



Definición 2 (Vector de residuos):

Un vector aleatorio m -dimensional $\mathbf{r}(\boldsymbol{\beta}; \mathbf{Y})$ es llamado un **vector de residuos** si:

- (i) \mathbf{r} es **insesgado**, es decir,

$$E_{\boldsymbol{\beta}}\{\mathbf{r}(\boldsymbol{\beta}; \mathbf{Y})\} = \mathbf{0}.$$

- (ii) \mathbf{r} tiene **matriz de variabilidad**

$$\mathbf{V}_r(\boldsymbol{\beta}) = \text{Cov}(\mathbf{r}(\boldsymbol{\beta}; \mathbf{Y}))$$

Observación:

- ▶ Un vector $\mathbf{r}(\boldsymbol{\beta})$ se dice **no-singular** si $\mathbf{V}_r(\boldsymbol{\beta})$ es definida positiva para todo $\boldsymbol{\beta}$.
- ▶ Note el énfasis que se hace sobre **supuestos de momentos**.



Un vector $\mathbf{r}(\beta)$ es llamado **suave** si $\mathbf{r}(\cdot; \mathbf{Y})$ es diferenciable para casi todo \mathbf{Y} , y la **matriz de sensibilidad** $m \times k$

$$\mathbf{S}_r(\beta) = \mathbb{E}_\beta \left\{ \frac{\partial \mathbf{r}(\beta)}{\partial \beta} \right\},$$

existe para todo β .

Definición 3 (Información de Godambe):

Para un vector residual no-singular y suave, $\mathbf{r}(\beta)$ se define la **matriz de información de Godambe**¹ como:

$$\mathbf{G}_r(\beta) = \mathbf{S}_r^\top(\beta) \mathbf{V}_r^{-1}(\beta) \mathbf{S}_r(\beta).$$

¹Note la forma de **sandwich**



Sea $\mathbf{A}(\beta)$ matriz $p \times m$ no estocástica y considere

$$\mathbf{q}(\beta) = \mathbf{A}(\beta)\mathbf{r}(\beta).$$

De ahí que

$$\mathbf{V}_q(\beta) = \text{Cov}(\mathbf{q}(\beta)) = \mathbf{A}(\beta)\mathbf{V}_r(\beta)\mathbf{A}^\top(\beta).$$

Resultado 1:

Para $\mathbf{A}(\beta)$ matriz $p \times m$ no estocástica y $\mathbf{r}(\beta)$ regular. Entonces

$$\mathbf{S}_q(\beta) = \mathbf{A}(\beta)\mathbf{S}_r(\beta).$$

Demostración:

En efecto,

$$\begin{aligned}\mathbf{S}_q(\beta) &= \mathbf{E}\{\nabla_\beta \mathbf{q}(\beta)\} = \mathbf{E}\{[\nabla_\beta \mathbf{A}(\beta)]\mathbf{r}(\beta) + \mathbf{A}(\beta)\nabla_\beta \mathbf{r}(\beta)\} \\ &= [\nabla_\beta \mathbf{A}(\beta)] \mathbf{E}\{\mathbf{r}(\beta)\} + \mathbf{A}(\beta) \mathbf{E}\{\nabla_\beta \mathbf{r}(\beta)\} \\ &= \mathbf{A}(\beta)\mathbf{S}_r(\beta)\end{aligned}$$



Considere $\mathbf{r}(\boldsymbol{\beta})$ y $\mathbf{s}(\boldsymbol{\beta})$ vectores residuales m -dimensionales. Entonces,

$$\mathbf{S}_{r+s}(\boldsymbol{\beta}) = \mathbf{S}_r(\boldsymbol{\beta}) + \mathbf{S}_s(\boldsymbol{\beta}).$$

La linealidad de $\mathbf{S}(\boldsymbol{\beta})$ es evidentemente, pues

$$\nabla_{\boldsymbol{\beta}}[\mathbf{r}(\boldsymbol{\beta}) + \mathbf{s}(\boldsymbol{\beta})] = \nabla_{\boldsymbol{\beta}}\mathbf{r}(\boldsymbol{\beta}) + \nabla_{\boldsymbol{\beta}}\mathbf{s}(\boldsymbol{\beta}).$$

Si además $\mathbf{r}(\boldsymbol{\beta})$ y $\mathbf{s}(\boldsymbol{\beta})$ son independientes. Entonces,

$$\mathbf{V}_{r+s}(\boldsymbol{\beta}) = \mathbf{V}_r(\boldsymbol{\beta}) + \mathbf{V}_s(\boldsymbol{\beta}).$$



Considere una re-parametrización $\beta = \beta(\gamma)$, donde γ es un vector l -dimensional ($l \leq k$), y sea

$$q(\gamma) = r(\beta(\gamma))$$

Entonces,

$$V_q(\gamma) = V_r(\beta(\gamma)), \quad S_q(\gamma) = S_r(\beta(\gamma)) \nabla_\gamma \beta(\gamma).$$

En efecto

$$\begin{aligned} S_q(\gamma) &= E_\gamma \{ \nabla_\gamma q(\gamma) \} = E_\gamma \{ \nabla_\gamma r(\beta(\gamma)) \} = E_\gamma \{ \nabla_\beta r(\beta(\gamma)) \nabla_\gamma \beta(\gamma) \} \\ &= E_\gamma \{ \nabla_\beta r(\beta) \} \nabla_\gamma \beta(\gamma) = S_r(\beta(\gamma)) \nabla_\gamma \beta(\gamma) \end{aligned}$$

Adicionalmente, si r es no-singular, entonces

$$\begin{aligned} G_q(\gamma) &= S_q^\top(\gamma) V_q^{-1}(\gamma) S_q(\gamma) \\ &= \nabla_\gamma^\top \beta(\gamma) S_r^\top(\gamma) V_r^{-1}(\gamma) S_r(\beta(\gamma)) \nabla_\gamma \beta(\gamma) \\ &= \nabla_\gamma^\top \beta(\gamma) G_r(\beta(\gamma)) \nabla_\gamma \beta(\gamma) \end{aligned}$$



Definición 4:

Un vector residual suave m -dimensional $\mathbf{r}(\beta)$ es dicho **regular**, si $k \leq m$ y la matriz de sensibilidad $\mathbf{S}_r(\beta)$ tiene rango k para todo β .

Resultado 2:

Un vector residual regular $\mathbf{r}(\beta)$ es no-singular.

Demostración:

(Por contradicción) Suponga que $\mathbf{r}(\beta)$ es singular, en cuyo caso $\mathbf{V}_r(\beta) = \text{Cov}(\mathbf{r}(\beta))$ no es definida positiva. Entonces, existe un vector no nulo \mathbf{c} tal que la combinación $\mathbf{c}^\top \mathbf{r}(\beta)$ es cero casi seguramente. De ahí que

$$\mathbf{0} = \mathbf{S}_{\mathbf{c}^\top \mathbf{r}}(\beta) = \mathbf{c}^\top \mathbf{S}_r(\beta),$$

y de ahí que $\mathbf{S}_r(\beta)$ es singular, lo que contradice que $\mathbf{r}(\beta)$ sea regular. De ahí que $\mathbf{V}_r(\beta)$ debe ser definida positiva.



Para un vector residual regular $\mathbf{r}(\beta)$, la matriz de información de Godambe

$$\mathbf{G}_r(\beta) = \mathbf{S}_r^\top(\beta) \mathbf{V}_r^{-1}(\beta) \mathbf{S}_r(\beta),$$

es definida positiva.

Para función de estimación regular $\Psi_n : \mathbb{R}^k \rightarrow \mathbb{R}^k$, la matriz de sensibilidad será cuadrada aunque no necesariamente simétrica.

Suponga $\Phi_n(\beta) = \mathbf{A}(\beta) \Psi_n(\beta)$ con $\mathbf{A}(\beta)$ matriz no singular $k \times k$. Entonces

$$\mathbf{G}_\Phi(\beta) = \mathbf{S}_\Psi^\top(\beta) \mathbf{A}(\beta) \{ \mathbf{A}^{-\top}(\beta) \mathbf{V}_\Psi^{-1}(\beta) \mathbf{A}^{-1}(\beta) \} \mathbf{A}(\beta) \mathbf{S}_\Psi(\beta)$$

$$\mathbf{S}_\Psi^\top(\beta) \mathbf{V}_\Psi^{-1}(\beta) \mathbf{S}_\Psi(\beta) = \mathbf{G}_\Psi(\beta),$$

de ahí que Φ y Ψ son equivalentes.



Definición 5 (Algoritmo Newton-scoring):

Considere la expansión en series de Taylor de $\Psi(\beta)$ en torno de β_0 , tenemos

$$\Psi(\beta) \approx \Psi(\beta_0) + \left. \frac{\partial \Psi(\beta)}{\partial \beta^\top} \right|_{\beta=\beta_0} (\beta - \beta_0),$$

como $\Psi(\hat{\beta}) = 0$ y substituyendo $\dot{\Psi}(\beta_0)$ por $S_\Psi(\beta_0)$, sigue que

$$\hat{\beta} = \beta_0 - S_\Psi^{-1}(\beta_0) \Psi(\beta_0), \quad (1)$$

esto sugiere considerar:

$$\beta^{(t+1)} = \beta^{(t)} - S_\Psi^{-1}(\beta^{(t)}) \Psi(\beta^{(t)}),$$

para llevar a cabo la estimación de parámetros.

Observación:

Este procedimiento fue propuesto por Jørgensen y Knudsen (2004)² quienes lo denominaron **algoritmo Newton-scoring**.

²Scandinavian Journal of Statistics **31**, 93-114.



Ejemplo:

Considere la función de estimación de mínimos cuadrados no lineales

$$\Psi_n(\beta) = \sum_{i=1}^n \mathbf{f}_i(\beta)(Y_i - f(\mathbf{x}_i; \beta)),$$

con $\mathbf{f}_i(\beta) = \partial f(\mathbf{x}_i; \beta) / \partial \beta$. En efecto, $E\{\Psi_n(\beta)\} = \mathbf{0}$, y

$$\begin{aligned} V(\beta) &= \sum_{i=1}^n \mathbf{f}_i(\beta) \text{var}(Y_i - f(\mathbf{x}_i; \beta)) \mathbf{f}_i^\top(\beta) = \sigma^2 \sum_{i=1}^n \mathbf{f}_i(\beta) \mathbf{f}_i^\top(\beta) \\ &= \sigma^2 \mathbf{F}^\top \mathbf{F}, \end{aligned}$$

donde $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_n)^\top$. Mientras que

$$S(\beta) = - \sum_{i=1}^n \mathbf{f}_i(\beta) \mathbf{f}_i^\top(\beta).$$

De este modo,

$$G(\beta) = \sigma^2 \mathbf{F}^\top \mathbf{F}.$$



Definición 6:

Sea $\mathbf{r}(\beta)$ vector de residuos no singular. Defina la **función quasi-score** como:

$$\mathbf{q}(\beta) = -\mathbf{S}_r^\top(\beta) \mathbf{V}_r^{-1}(\beta) \mathbf{r}(\beta),$$

donde \mathbf{S}_r y \mathbf{V}_r son las matrices de sensibilidad y variabilidad de $\mathbf{r}(\beta)$.

La matriz $k \times m$,

$$\mathbf{A}(\beta) = -\mathbf{S}_r^\top(\beta) \mathbf{V}_r^{-1}(\beta),$$

es llamada **matriz de pesos de Crowder** (1987).³ Además, las matrices de sensibilidad y variabilidad de $\mathbf{q}(\beta)$ son

$$\mathbf{S}_q = -\mathbf{S}_r^\top \mathbf{V}_r^{-1} \mathbf{S}_r = -\mathbf{G}_r$$

$$\mathbf{V}_q = \mathbf{S}_r^\top \mathbf{V}_r^{-1} \mathbf{V}_r \mathbf{V}_r^{-1} \mathbf{S}_r = \mathbf{G}_r.$$

Es decir, $\mathbf{V}_q = -\mathbf{S}_q$, y de ahí que

$$\mathbf{G}_q = \mathbf{G}_r.$$

³Biometrika 74, 591-597.



Observación:

Cualquier función de estimación regular $\Psi(\beta)$ satisfaciendo

$$-S_{\Psi}(\beta) = V_{\Psi}(\beta),$$

es una función quasi-score.

En particular, la **función score**

$$U(\beta) = \frac{\partial}{\partial \beta} \log p(\beta),$$

es una función quasi-score.



Suponga $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ vectores aleatorios independientes y sea

$$\mathbf{r}_i(\boldsymbol{\mu}_i), \quad \boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}),$$

vectores m_i -dimensionales. Considere la siguiente definición.

Definición 7:

Una **función de estimación lineal** es definida como:

$$\mathbf{g}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{W}_i(\boldsymbol{\beta}) \mathbf{r}_i(\boldsymbol{\mu}_i),$$

donde $\mathbf{W}_i = \mathbf{W}_i(\boldsymbol{\beta})$ es una matriz no aleatoria $k \times m_i$.



Ejemplo (mínimos cuadrados):

En este caso tenemos

$$\mathbf{g}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}).$$

Ejemplo (M-estimación):

Considere Y_i con distribución simétrica, la función de estimación M es

$$\mathbf{g}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}),$$

donde ψ es una función impar.



Ejemplo (GLM):

Suponga que

$$\mathbf{g}(\boldsymbol{\beta}) = \phi \sum_{i=1}^n \left(\frac{\omega_i}{V_i} \right)^{1/2} \mathbf{x}_i (Y_i - \mu_i),$$

donde $E(Y_i) = \mu_i$, con $h(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$.

Ejemplo (GEE):

Considere

$$\mathbf{g}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{D}_i^\top \mathbf{C}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i),$$

donde \mathbf{Y}_i es un vector aleatorio con media $\boldsymbol{\mu}_i$, $\mathbf{D}_i = \nabla_{\boldsymbol{\beta}} \boldsymbol{\mu}_i$ es la matriz de modelo local y \mathbf{C}_i es la matriz de covarianza de trabajo.



Previo (Ordenamiento de Löwner):

Si A y B son matrices de covarianza de la misma dimensión. Entonces $A \geq B$ si y sólo si $A - B$ es semidefinida positiva.⁴

Definición 8:

Basado en la Ecuación (1) podemos definir una **función de estimación normalizada** como:

$$\bar{g}(\beta) = -S_g^{-1}(\beta)g(\beta).$$

⁴Esto corresponde a un ordenamiento parcial en el espacio de matrices semidefinidas positivas.



Resultado 3:

Sea \bar{g}_1 y \bar{g}_2 dos funciones de estimación normalizadas con matrices de información de Godambe K_1 y K_2 , respectivamente. La condición

$$\text{Cov}(\bar{g}_1, \bar{g}_2) = K_1,$$

implica las siguientes tres condiciones:

- (a) \bar{g}_1 y $\bar{g}_2 - \bar{g}_1$ son no correlacionadas.
- (b) $K_2 \geq K_1$.
- (c) $\text{var}(c^\top \bar{g}_2) \geq \text{var}(c^\top \bar{g}_1)$ para cualquier vector $c \in \mathbb{R}^k$.



Demostración:

Note que podemos escribir

$$\bar{\mathbf{g}}_2 = \bar{\mathbf{g}}_1 + (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1). \quad (2)$$

Si $\text{Cov}(\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_2) = \mathbf{K}_1$ entonces

$$\begin{aligned} \text{Cov}(\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1) &= \text{Cov}(\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_2) - \text{Cov}(\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_1) \\ &= \text{Cov}(\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_2) - \mathbf{K}_1 = \mathbf{0}, \end{aligned}$$

lo que prueba (a). Tomando covarianzas en ambos lados de (2) lleva a

$$\mathbf{K}_2 = \mathbf{K}_1 + \mathbf{A}, \quad \mathbf{A} = \text{Cov}(\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1),$$

como \mathbf{A} es semidefinida positiva, sigue (b). Además, tenemos que para cualquier \mathbf{c} ,

$$\begin{aligned} 0 \leq \mathbf{c}^\top \mathbf{A} \mathbf{c} &= \mathbf{c}^\top (\mathbf{K}_2 - \mathbf{K}_1) \mathbf{c} = \mathbf{c}^\top \mathbf{K}_2 \mathbf{c} - \mathbf{c}^\top \mathbf{K}_1 \mathbf{c} \\ &= \text{var}(\mathbf{c}^\top \bar{\mathbf{g}}_2) - \text{var}(\mathbf{c}^\top \bar{\mathbf{g}}_1), \end{aligned}$$

lo que implica (c).



Resultado 4 (Optimalidad de Crowder):

Suponga $q(\beta)$ función quasi-score

$$q(\beta) = \sum_{i=1}^n W_i(\beta) r_i(\mu_i).$$

Entonces $q(\beta)$ es óptima para todas las funciones de estimación lineales en el sentido que

$$G_g^{-1}(\beta) \geq G_q^{-1}(\beta),$$

para cualquier función de estimación lineal $g(\beta)$, con la igualdad si y sólo si

$$\bar{g}(\beta) = \bar{q}(\beta),$$

esto es, si $g(\beta)$ y $q(\beta)$ son equivalentes.

