MAT-466: Propiedades de funciones de inferencia

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Resultado 1:

Sea $\overline{\boldsymbol{g}}_1$ y $\overline{\boldsymbol{g}}_2$ dos funciones de estimación normalizadas con matrices de información de Godambe $\boldsymbol{K}_1$ y $\boldsymbol{K}_2$, respectivamente. La condición

$$\mathsf{Cov}(\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_2) = \boldsymbol{K}_1, \tag{1}$$

implica las siguientes tres condiciones:

(a) $\overline{\boldsymbol{g}}_1$ y $\overline{\boldsymbol{g}}_2 - \overline{\boldsymbol{g}}_1$ son no correlacionadas.

(b) $\boldsymbol{K}_2 \geq \boldsymbol{K}_1$.

(c) $\mathsf{var}(\boldsymbol{c}^\top \overline{\boldsymbol{g}}_2) \geq \mathsf{var}(\boldsymbol{c}^\top \overline{\boldsymbol{g}}_1)$ para cualquier vector $\boldsymbol{c} \in \mathbb{R}^k$.

Demostración:

Note que podemos escribir

$$\overline{\boldsymbol{g}}_2 = \overline{\boldsymbol{g}}_1 + (\overline{\boldsymbol{g}}_2 - \overline{\boldsymbol{g}}_1). \qquad (2)$$

Si $\mathsf{Cov}(\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_2) = \boldsymbol{K}_1$ entonces

$$\begin{aligned}\mathsf{Cov}(\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_2 - \overline{\boldsymbol{g}}_1) &= \mathsf{Cov}(\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_2) - \mathsf{Cov}(\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_1) \\ &= \mathsf{Cov}(\overline{\boldsymbol{g}}_1, \overline{\boldsymbol{g}}_2) - \boldsymbol{K}_1 = \boldsymbol{0},\end{aligned}$$

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

$$\boldsymbol{K}_2 = \boldsymbol{K}_1 + \boldsymbol{A}, \qquad \boldsymbol{A} = \mathsf{Cov}(\overline{\boldsymbol{g}}_2 - \overline{\boldsymbol{g}}_1),$$

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

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

lo que implica (c).

Considere la función de inferencia

$$\boldsymbol{q}(\boldsymbol{\beta}) = -\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{r}_i(\boldsymbol{\mu}_i), \qquad \boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}^\top.$$

Note que

$$\boldsymbol{S}_q = -\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{S}_{r_i} = -\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{S}_i \boldsymbol{D}_i,$$

y

$$\begin{aligned}\boldsymbol{V}_q &= \sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{V}_i \boldsymbol{V}_i^{-1} \boldsymbol{S}_i \boldsymbol{D}_i = \sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{S}_i \boldsymbol{D}_i \\ &= -\boldsymbol{S}_q.\end{aligned}$$

De este modo, $\boldsymbol{q}(\boldsymbol{\beta})$ es una función quasi-score con información de Godambe

$$\boldsymbol{G}_q(\boldsymbol{\beta}) = -\boldsymbol{S}_q = \boldsymbol{V}_q.$$

Por otro lado, suponga que

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

De este modo,

$$\boldsymbol{S}_g = \sum_{i=1}^{n} \boldsymbol{W}_i \boldsymbol{S}_{r_i} = \sum_{i=1}^{n} \boldsymbol{W}_i \boldsymbol{S}_i \boldsymbol{D}_i,$$

mientras que

$$\boldsymbol{V}_g = \sum_{i=1}^{n} \mathsf{Cov}\left(\boldsymbol{W}_i(\boldsymbol{\beta})\boldsymbol{r}_i(\boldsymbol{\mu}_i)\right) = \sum_{i=1}^{n} \boldsymbol{W}_i \boldsymbol{V}_i \boldsymbol{W}_i^{\top}.$$

Es decir,

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

Resultado 2 (Optimalidad de Crowder):

Suponga $\boldsymbol{q}(\boldsymbol{\beta})$ función quasi-score

$$\boldsymbol{q}(\boldsymbol{\beta}) = -\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{r}_i(\boldsymbol{\mu}_i)$$

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

$$\boldsymbol{G}_g^{-1}(\boldsymbol{\beta}) \geq \boldsymbol{G}_q^{-1}(\boldsymbol{\beta}),$$

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

$$\overline{\boldsymbol{g}}(\boldsymbol{\beta}) = \overline{\boldsymbol{q}}(\boldsymbol{\beta}),$$

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

Demostración:

Deseamos mostrar la Ecuación (1) del Resultado 1. En efecto

$$
\begin{aligned}
\mathsf{Cov}(\boldsymbol{q},\boldsymbol{g}) = \mathsf{E}(\boldsymbol{q}\boldsymbol{g}^\top) &= -\,\mathsf{E}\Big[\sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{r}_i\Big(\sum_{j=1}^n \boldsymbol{W}_j\boldsymbol{r}_j\Big)\Big] \\
&= -\,\mathsf{E}\Big[\sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{r}_i \sum_{j=1}^n \boldsymbol{r}_j^\top \boldsymbol{W}_j^\top\Big] \\
&= -\sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1} \sum_{j=1}^n \mathsf{E}(\boldsymbol{r}_i\boldsymbol{r}_j^\top)\boldsymbol{W}_j^\top,
\end{aligned}
$$

como $\boldsymbol{r}_i$ y $\boldsymbol{r}_j$ son no correlacionados para $i \neq j$, sigue que

$$
\begin{aligned}
\mathsf{E}(\boldsymbol{q}\boldsymbol{g}^\top) &= -\sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1}\,\mathsf{E}(\boldsymbol{r}_i\boldsymbol{r}_i^\top)\boldsymbol{W}_j^\top = -\sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{V}_i\boldsymbol{W}_j^\top \\
&= -\sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{S}_i^\top \boldsymbol{W}_i^\top = -\boldsymbol{S}_g^\top
\end{aligned}
$$

Esto permite calcular

$$\begin{aligned}\mathsf{Cov}(\overline{\boldsymbol{q}}, \overline{\boldsymbol{g}}) &= \mathsf{E}\{\boldsymbol{S}_q^{-1}\boldsymbol{q}(\boldsymbol{S}_g^{-1}\boldsymbol{g})^\top\} = \boldsymbol{S}_q^{-1}\,\mathsf{E}(\boldsymbol{q}\boldsymbol{g}^\top)\boldsymbol{S}_g^{-\top} \\ &= -\boldsymbol{S}_q^{-1}\boldsymbol{S}_g^\top\boldsymbol{S}_g^{-\top} = -\boldsymbol{S}_q^{-1} = \boldsymbol{G}_q^{-1},\end{aligned}$$

y esto implica la optimalidad de $\boldsymbol{q}$, lo que concluye la prueba.

Asuma una función de estimación lineal

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

donde la notación enfatiza la dependencia de n. Asimismo $\boldsymbol{S}_n(\boldsymbol{\beta})$ y $\boldsymbol{V}_n(\boldsymbol{\beta})$ denotan las matrices de sensibilidad y variabilidad de $\boldsymbol{g}_n$.

Además, suponga que

$$\lim_{n\to\infty} \frac{1}{n}\boldsymbol{S}_n(\boldsymbol{\beta}) = \boldsymbol{S}(\boldsymbol{\beta}), \qquad \lim_{n\to\infty} \frac{1}{n}\boldsymbol{V}_n(\boldsymbol{\beta}) = \boldsymbol{V}(\boldsymbol{\beta}).$$

Resultado 3:

Si $\boldsymbol{S}(\boldsymbol{\beta})$ es no singular, entonces $\boldsymbol{V}(\boldsymbol{\beta})$ es no singular.

Demostración:

La prueba se lleva a cabo por contradicción. Suponga que $\boldsymbol{V}(\boldsymbol{\beta})$ es singular. Entonces existe un vector $\boldsymbol{c}$ tal que

$$\boldsymbol{c}^\top \boldsymbol{V}(\boldsymbol{\beta})\boldsymbol{c} = 0,$$

y de ahí que

$$\lim_{n\to\infty} \frac{1}{n}\boldsymbol{c}^\top \boldsymbol{V}_n(\boldsymbol{\beta})\boldsymbol{c} = 0,$$

lo cual implica que

$$\frac{1}{n}\boldsymbol{c}^\top \boldsymbol{g}_n(\boldsymbol{\beta}) \xrightarrow{\mathsf{p}} 0.$$

Asumiendo la continuidad del operador de sensibilidad $\boldsymbol{S}(\cdot)$, tenemos

$$\begin{aligned}
\boldsymbol{0} &= \boldsymbol{S}\Big[\lim_{n\to\infty} \boldsymbol{c}^\top \boldsymbol{g}_n(\boldsymbol{\beta})\Big] = \lim_{n\to\infty} \boldsymbol{S}\Big[\frac{1}{n}\boldsymbol{c}^\top \boldsymbol{g}_n(\boldsymbol{\beta})\Big] \\
&= \boldsymbol{c}^\top \lim_{n\to\infty} \frac{1}{n}\boldsymbol{S}_n(\boldsymbol{\beta}) = \boldsymbol{c}^\top \boldsymbol{S}(\boldsymbol{\beta})
\end{aligned}$$

lo que es una contradicción pues $\boldsymbol{S}(\boldsymbol{\beta})$ ha sido asumida no singular.

Note que

$$\frac{1}{n}\boldsymbol{G}_n(\boldsymbol{\beta}) = \frac{1}{n}\boldsymbol{S}_n^\top(\boldsymbol{\beta})\Big[\frac{1}{n}\boldsymbol{V}_n(\boldsymbol{\beta})\Big]^{-1}\frac{1}{n}\boldsymbol{S}_n(\boldsymbol{\beta}) \\ \to \boldsymbol{S}^\top(\boldsymbol{\beta})\boldsymbol{V}^{-1}(\boldsymbol{\beta})\boldsymbol{S}(\boldsymbol{\beta}) = \boldsymbol{G}(\boldsymbol{\beta}),$$

es decir, la información de Godambe promedio converge a un límite finito.

Asuma también que (Yuan y Jennrich, 1998)[1]

$$\frac{1}{\sqrt{n}}\boldsymbol{g}_n(\boldsymbol{\beta}) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{V}(\boldsymbol{\beta})).$$

Usando que

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \overset{\mathsf{a}}{=} \sqrt{n}\overline{\boldsymbol{g}}_n(\boldsymbol{\beta}) = -\Big[\frac{1}{n}\boldsymbol{S}_n(\boldsymbol{\beta})\Big]^{-1}\frac{1}{\sqrt{n}}\boldsymbol{g}_n(\boldsymbol{\beta}) \\ \xrightarrow{\mathsf{D}} \mathsf{N}_k\big(\mathbf{0}, \boldsymbol{S}^{-1}(\boldsymbol{\beta})\boldsymbol{V}(\boldsymbol{\beta})\boldsymbol{S}^{-\top}(\boldsymbol{\beta})\big),$$

es decir

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathsf{D}} \mathsf{N}_k\big(\mathbf{0}, \boldsymbol{G}^{-1}(\boldsymbol{\beta})\big).$$

[1] Journal of Multivariate Analysis 65, 245-260.

Test de hipótesis en funciones de inferencia

Problema:

Se desea probar hipótesis de la forma:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0, \qquad \text{versus} \qquad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,$$

donde $\boldsymbol{\beta}_0$ es un valor fijado.

Consideraremos los siguientes estadísticos de prueba:

- Test de Wald.
- Test tipo-score (de multiplicadores de Lagrange).

Observación:

- En el contexto de funciones de inferencia este tipo de test han sido discutidos por Rotnitzky y Jewell (1990)[2] y Boos (1992)[3]
- Mientras que Crudu y Osorio (2020)[4] abordan el test gradiente (de forma bilineal) en extremum estimation.

[2] Biometrika 77, 485-497.
[3] The American Statistician 46, 327-333.
[4] Economic Letters 187, 108885

Test de hipótesis en funciones de inferencia

Basado en la normalidad asintótica del estimador $\widehat{\boldsymbol{\beta}}_n$, sigue que

$$\boldsymbol{Z}_n = \sqrt{n}\boldsymbol{G}^{1/2}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\boldsymbol{0}, \boldsymbol{I}).$$

de ahí que

$$\begin{aligned} W_n = \boldsymbol{Z}_n^\top \boldsymbol{Z}_n &= n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^\top \boldsymbol{G}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\ &\xrightarrow{\mathsf{D}} \chi^2(k). \end{aligned}$$

Lo que lleva al test de tipo-Wald para probar $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, que es definido por la región crítica

$$\{W_n \geq \chi^2_{a-\alpha}(k)\}$$

que bajo H_0 es asintóticamente de tamaño α.

EX UMBRA IN SOLEM

Por otro lado, usando la condición

$$\frac{1}{\sqrt{n}}\boldsymbol{g}_n(\boldsymbol{\beta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{V}(\boldsymbol{\beta}_0)),$$

así

$$\boldsymbol{V}^{-1/2}(\boldsymbol{\beta}_0)\frac{1}{\sqrt{n}}\boldsymbol{g}_n(\boldsymbol{\beta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{I}),$$

lo que permite definir el estadístico tipo-score para probar hipótesis de la forma $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, como:

$$R_n = \frac{1}{n}\boldsymbol{g}_n^\top(\boldsymbol{\beta})\boldsymbol{V}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{g}_n(\boldsymbol{\beta}_0) \xrightarrow{\mathsf{D}} \chi^2(k).$$

Observación:

- Lamentablemente no es simple extender los estadísticos tipo-Wald y tipo-score para manipular hipótesis más generales, por ejemplo:

$$H_0^1 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}, \qquad H_0^2 : \boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}, \qquad H_0^3 : \boldsymbol{\beta} = \boldsymbol{h}(\boldsymbol{\lambda}),$$

  que son conocidos como hipótesis sobre sub-vectores y hipótesis no lineales en forma implícita y explícita, respectivamente.

- Algún esfuerzo se está llevando a cabo en extender el test gradiente (o de forma bilineal) a este contexto más general (ver, por ejemplo, Crudu y Osorio, 2020).