Elementos de Teoria Asintotica

Walter Sosa Escudero
(wsosa@udesa.edu.ar)

Universidad de San Andres

El modelo lineal en notacion observacional

$$y_i = x_i'\beta + u_i, \qquad i = 1, 2, \ldots, n$$

$$x_i = \begin{bmatrix} M_i \\ Z_i \end{bmatrix}, \quad x_i x_i' = \begin{bmatrix} M_i \\ Z_i \end{bmatrix} \begin{bmatrix} M_i & Z_i \end{bmatrix} = \begin{bmatrix} M_i^2 & M_i Z_i \\ Z_i M_i & Z_i^2 \end{bmatrix}$$

$$\sum_{i=1}^{n} x_i x_i' = \sum_{i=1}^{n} \begin{bmatrix} M_i^2 & M_i Z_i \\ Z_i M_i & Z_i^2 \end{bmatrix} = X'X$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \begin{bmatrix} M_i y_i \\ Z_i y_i \end{bmatrix} = X'Y$$

Entonces, en notacion observacional:

$$\hat{\beta} = (X'X)^{-1}X'Y = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} x_i y_i\right)$$

Propiedades Asintoticas, o de Muestra Grande

- Propiedades de muestra finita: propiedades para un tamaño de muestra fijo (insesgadez, varianza minima, etc.).
- Propiedades de muestra grande: propiedades de un estimador cuando el tamaño de la muestra se hace infinitamente grande.

- Porque?: en general son mas faciles de verificar que las de muestras finitas.
- Relevancia: en la practica funcionan correctamente aun cuando el tamaño de muestra no es infinito.

Convergencia de sucesiones aleatorias

- Concepto 'primitivo': $a_n \to a_0$, a_n es una sucesion de numeros reales.

  Ejemplo: $a_n = \frac{1}{n} \to 0$

- La idea es extender esta nocion de convergencia cuando hablamos de una sucesion de variables aleatorias.

Convergencia en probabilidad

x_n, sucesion de variables aleatorias. x_n converge en probabilidad a una constante a si para todo $\epsilon > 0$

$$Pr\left[|x_n - a| > \epsilon\right] \to 0$$

Usaremos la notacion $x_n \xrightarrow{p} a$, o *plim* $x_n = a$

- ver grafico e intuicion -

La idea se extiende a vectores de variables aleatorias, elemento por elemento.

Resultado importante:
Continuidad del *plim:* Si $x_n \xrightarrow{p} a$, y $g(z)$ es continua en a, entonces:

$$x_n \xrightarrow{p} a \Longrightarrow g(x_n) \xrightarrow{p} g(a)$$

Este resutado es muy importante: el plim 'pasa' a traves de funciones continuas.
Las esperanzas *no* tienen esta propiedad. Se generaliza a vectores.

- ¿Que era la 'ley de grandes numeros'?

El resultado central de convergencia en probabilidad es el siguiente:

Ley de Grandes Numeros

Si z_n es una sucesion de variables aleatorias independientes e identicamente distribuidas (iid) tal que $E(z_i) = \mu < \infty$ y $V(z_i) = \sigma^2 < \infty$, sea $\bar{z} \equiv (1/n)\sum_{i=1}^{n} z_i$ entonces $\bar{z}_n \xrightarrow{p} \mu$.

Consideremos una sucesion de estimadores $\hat{\theta}_n$, que se forman incrementando progresivamente el tamaño de la muestra.
Un estimador $\hat{\theta}_n$ es consistente para θ_0 si:

$$\hat{\theta}_n \xrightarrow{p} \theta_0$$

- Que dice la LGN en terminos de la propiedad de consistencia?

Convergencia en distribucion

x_n converge en distribucion a una variable aleatoria x sii:

$$F_n(\delta) \longrightarrow F(\delta)$$

para (casi) todo δ. Notacion: $x_n \xrightarrow{d} x$.
Estamos asociando la convergencia de una sucesion de variables aleatorias a la convergencia de sus funciones de distribucion.

- Que el Teorema Central del Limite?

El resultado central de convergencia en distribucion es:

Teorema Central de Limite (Lindeberg-Levy)

Sea z_i una sucesion de variables aleatorias iid con $E(z_i) = \mu < \infty$ y $V(u_i) = \sigma^2 < \infty$. Sea $\bar{z}_n \equiv n^{-1} \sum_{i=1}^{n} z_i$. Entonces:

$$\frac{\bar{z}_n - \mu}{\sigma/\sqrt{n}} \quad \xrightarrow{d} \quad N(0,1)$$

o, equivalentemente,

$$\sqrt{n}\,(\bar{z}_n - \mu) \quad \xrightarrow{d} \quad N(0, \sigma^2)$$

Es un resultado MUY importante

- Notar que habiendo establecido la consistencia de $\bar{z}_n$, la distribucion asintotica de $\bar{z}_n$ es trivial: porque?
- La confusion aparece por la siguiente aproximacion. Si $x \sim N(\mu, \sigma^2)$, entonces $a + bX \sim N(a + b\mu, b^2\sigma^2)$. Entonces, estamos tentados a decir que:

$$\sqrt{n}\,(\bar{z}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \Longrightarrow \bar{z}_n \sim N(\mu, \sigma^2/n)$$

pero cuando $n \to \infty$ el lado derecho es trivial!!!

$\bar{z}_n \sim N(\mu, \sigma^2/n)$ es solo una aproximacion.

Porque este resultado es tan importante?

- Consideremos el problema de evaluar $H_0 : \mu - \mu_0 = 0$ vs $H_0 : \mu - \mu_0 \neq 0$, chequeando si $\bar{x} - \mu_0 = 0$.
- Notar que siempre que $\mu - \mu_0 = 0$, $\delta(\mu - \mu_0) = 0$ para todo $\delta > 0$.
- Entonces, para evaluar H_0 podriamos mirar $\delta(\bar{x} - \mu_0) = 0$ Porque? Si elegimos inteligentemente δ, la distribucion de $\delta(\bar{x} - \mu_0)$ se puede derivar mas facilmente que la de $(\bar{x} - \mu_0)$. Por ejemplo, $\delta = \sqrt{n}/\sigma$, el TCL provee la distribucion de interes.

Dos conceptos importantes:

- Sea $\hat{\theta}_n$ una sucesion de estimadores. Si

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \sigma^2)$$

  diremos que $\hat{\theta}_n$ es asintoticamente normal con *varianza asintotica* σ^2/n. Diremos que el estimador es *consistente a la tasa* $\sqrt{n}$.
- Sean $\hat{\theta}_n$ y $\bar{\theta}_n$ dos estimadores consistentes a la tasa $\sqrt{n}$, con varianzas asintoticas V_1 y V_2. Entonces $\hat{\theta}_n$ es asintoticamente eficiente en relacion a $\bar{\theta}_n$ si $V_2 - V_1 \geq 0$.

Resultados utiles

- Si X y Y son VA independientes y $g(z)$ es continua, entonces $g(X)$ y $g(Z)$ son independientes.
- $V(y) = E(V(y|x)) + V(E(y|x))$
- Teorema de Slutzky: Si $Y_n \xrightarrow{d} Y$ y $X_n \xrightarrow{p} a$ y $Z_n \xrightarrow{p} b$, entonces:

$$X_n + Z_n Y_n \xrightarrow{d} a + bY$$

- Cramer-Wald Device: x_n una sucesion de vectores $K \times 1$.

$$x_n \xrightarrow{d} x \quad \Longleftrightarrow \quad c'x_n \xrightarrow{d} c'x$$

para todo vector $c \in \Re^K$. Este resultado facilita establecer la convergencia en distribucion de vectores, reduciendo el problema a combinaciones lineales arbitrarias.

- Continuous mapping theorem: Si $g(z)$ es continua:

$$x_n \xrightarrow{d} a \Longrightarrow g(x_n) \xrightarrow{d} g(a)$$

- Distribucion normal multivariada: x un vector de K VA's, tiene distribucion normal multivariada con media μ y varianza Ω si la funcion de densidad conjunta $f(x)$ es:

$$f(x) = \frac{1}{(2\pi)^{K/2}|\Omega|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu)\right\}$$

Denotamos $x \sim N(\mu, \Omega)$. Una propiedad importante es la de *linealidad*:

$$x \sim N(\mu, \Omega) \Rightarrow a + b'x \sim N(a + b'\mu, b'\Omega b)$$

para cualquier par de vectores a y b.

- Aplicacion importante del teorema de Cramer-Wald bajo normalidad: x_n un vector de K VA's, $c \in \Re^K$:

$$x_n \xrightarrow{d} N(\mu, \Omega) \iff c' x_n \xrightarrow{d} N(c'\mu, c'\Omega c)$$

Resultados asintoticos para el estimador MCO

$$y_i = x_i'\beta_0 + u_i, \quad i = 1, 2, \ldots, n$$

x_i es un vector de K variables explicativas.
Supondremos:

- $\rho(X) = K$.
- $E(u_i|x_i) = 0$.
- $V(u_i|x_i) = \sigma^2, \;\; E(u_i u_j|x_i) = 0, \;\; i \neq j$.
- $(y_i, x_i) \sim$ iid, con momentos hasta el 4to.
- $E(x_i x_i') = D$, finita y no-singular.

Consistencia

Resultado: $\hat{\beta}_n$ es consistente

$$\begin{aligned}
\hat{\beta}_n &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(X\beta + u) \\
&= \beta_0 + (X'X)^{-1}X'u \\
&= \beta_0 + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{n}\right)
\end{aligned}$$

Si podemos probar que $\left(\frac{X'X}{n}\right)^{-1}$ *no tiende a infinito, ¿que queda por probar para establecer consistencia?*

El elemento h del vector $\left(\frac{X'u}{n}\right)$ es:

$$\frac{1}{n}\sum_{i=1}^{n} x_{hi}u_i = \frac{1}{n}\sum_{i=1}^{n} z_i, \qquad z_i \equiv x_{hi}u_i$$

Notar que las z_i son variables aleatorias iid con:

$$\begin{aligned} E(z_i) &= E(x_{hi}u_i) = E(E(x_{hi}u_i|x_{hi})) = E(x_{hi}E(u_i|x_{hi})) = 0 \\ V(z_i) &= E(V(z_i|x_{hi})) + V(E(z_i|x_{hi})) \\ &= E(V(z_i|x_{hi}) = E(V(x_{hi}u_i|x_{hi})) = \sigma^2 E(x_{hi}^2) < \infty \end{aligned}$$

¿Para que hemos hecho esto ultimo?

Entonces, por la ley de grandes numeros:

$$\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{\sum_{i=1}^{n} x_{hi}u_i}{n} \quad \xrightarrow{p} \quad E(x_{hi}u_i) = 0$$

de modo que: $\left(\frac{X'u}{n}\right) \xrightarrow{p} 0$.

El elemento (h, j) de $\left(\frac{X'X}{n}\right)^{-1}$ es:

$$\frac{\sum_{i=1}^{n} x_{hi}x_{ji}}{n} \xrightarrow{p} E(x_{hi}x_{ji}) = D_{hj}$$

Entonces:

$$\frac{X'X}{n} \xrightarrow{p} D$$

Por continuidad de plim:

$$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{p} D^{-1}$$

Resumiendo:

$$\hat{\beta}_n = \beta_0 + \underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\xrightarrow{p} D^{-1}} \underbrace{\left(\frac{X'u}{n}\right)}_{\xrightarrow{p} 0}$$

Entonces, por continuidad: $\hat{\beta}_n \xrightarrow{p} \beta_0$.

¿Cual es el supuesto 'mas importante' que garantiza la consistencia?

Normalidad asintotica

Resultado: $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{d} N(0, \sigma^2 D^{-1})$.

Es un poquito mas delicado. El punto de partida es, nuevamente:

$$\hat{\beta}_n = \beta_0 + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{n}\right)$$

Multipliquemos ambos lados por $\sqrt{n}$ y restemos β_0:

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{\sqrt{n}}\right)$$

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{\sqrt{n}}\right)$$

Ya mostramos que:

$$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{p} D^{-1}$$

La idea es buscar la distribucion asintotica de $\left(\frac{X'u}{\sqrt{n}}\right)$.

Si probamos que $\left(\frac{X'u}{\sqrt{n}}\right)$ *es asintoticamente normal, ¿como termina la prueba?*

Comencemos con:

$$\left(\frac{X'u}{\sqrt{n}}\right) = \sqrt{n}\frac{X'u}{n}$$

Es un vector de K VA's. Por el Cramer-Wold Device, es equivalente buscar la distribucion de:

$$\sqrt{n}c'\left(\frac{X'u}{n}\right)$$

para todo $c \in \Re^K$. En notacion observacional:

$$\sqrt{n}c'\left(\frac{X'u}{n}\right) = \sqrt{n}c'\frac{\sum x_i u_i}{n} = \sqrt{n}\frac{\sum c' x_i u_i}{n} \equiv \sqrt{n}\frac{\sum z_i}{n}$$

donde $z_i \equiv c' x_i u_i$. (probar como ejercicio!!)

Ahora es facil establecer las siguientes propiedades:

- $E(z_i) = 0$
- $V(z_i) = \sigma^2 c' D c < \infty$

Entonces, por el TCL aplicado a $\sqrt{n}\bar{z}$:

$$\sqrt{n}(\bar{z} - 0) = c' \left(\frac{X'u}{\sqrt{n}} \right) \xrightarrow{d} N(0, \sigma^2 c' D c)$$

Entonces, por el teorema de Cramer-Wald:

$$\left(\frac{X'u}{\sqrt{n}} \right) \xrightarrow{d} N(0, \sigma^2 D)$$

Reemplazando:

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = \underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\xrightarrow{p} D^{-1}} \quad \underbrace{\left(\frac{X'u}{\sqrt{n}}\right)}_{\xrightarrow{d} N(0,\sigma^2 D)}$$

Por Teorema de Slutzky y propiedad de linealidad de la distribucion normal multivariada:

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{p} N(0, \sigma^2 D^{-1} D D^{-1}) = N(0, \sigma^2 D^{-1})$$