CONCEPTOS BASICOS DE TEORIA ASINTOTICA

1. Convergencia en Probabilidad

La variable aleatoria x_n converge en probabilidad a una constante c si

$$\lim_{n\to\infty} \text{Prob}(|x_n - c| > \varepsilon) = 0$$

para cualquier $\varepsilon>0$.

La definición anterior indica que se hace cada vez más improbable que x_n tome valores distintos a c, a medida que n, el tamaño de la muestra, aumenta.

Ejemplo 1

Supongamos que tenemos una variable aleatoria x_n cuya distribución de probabilidad es la siguiente:

$$f(x_n) = \begin{cases} 1 - \frac{1}{n} & \text{si } x_n = 0 \\ \frac{1}{n} & \text{si } x_n = n \end{cases}$$

En este caso,

$$\lim_{n\to\infty} \text{Prob}(|x_n - 0| > \varepsilon) = 0$$

Es decir, x_n converge en probabilidad a 0. ¿Por qué? La razón es que, a medida que n aumenta, x_n toma el valor de n con una probabilidad cada vez menor (1/n converge a 0 a medida que $n\to\infty$). Esto es, toda la masa de la distribución se concentra en aquellos puntos en la vecindad de 0♦

En general, si x_n converge en probabilidad a c, escribimos

$$\text{plim } x_n = c \qquad \text{o} \qquad x_n \xrightarrow{p} c$$

La convergencia en probabilidad se denomina **convergencia débil**. En contraste, la convergencia “casi segura” (*almost surely* o “a.s”) o con probabilidad 1 se denomina **convergencia fuerte**. Esta se define de la siguiente manera:

$$\text{Prob}\{\omega | \lim_{n\to\infty} x_n(\omega)=x(\omega)\} = 1$$

Esto es, la secuencia $\{x_n\}$ converge a x con probabilidad 1. Esto se simboliza

$$x_n \xrightarrow{a.s} x$$

Ejemplos de convergencia fuerte son los siguientes:

i) Si $\{x_n\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas con $E(x_n)=\mu<\infty$, entonces:

$$\bar{x}_n \xrightarrow{a.s} \mu$$

por la ley fuerte de los grandes números.

ii) $\text{Prob}\{\lim_{n\to\infty} x_n=0\} = 1 \Leftrightarrow x_n \xrightarrow{a.s} 0$.

<u>Nota:</u> Es común encontrar en libros avanzados de econometría las notaciones O(1/n) y o(1/n). Se dice que c_n es O(1/n) si plim (nc_n) es una constante finita distinta de cero. En tanto, se dice que c_n es o(1/n) si plim $(nc_n)=0$. Por ejemplo,

$$c_n = \frac{1}{n} + \frac{3}{n^2} \text{ es O(1/n), dado que plim}(nc_n)=1$$

$$c_n = \frac{1}{n^2} \text{ es o(1/n), dado que plim}(nc_n)=0$$

2.- <u>Convergencia en Media Cuadrática</u>

Si x_n tiene media μ_n y varianza σ_n^2, tal que:

$$\lim_{n\to\infty} \mu_n = \mu \qquad \lim_{n\to\infty}\sigma_n^2=0$$

entonces decimos que x_n converge en media cuadrática (*quadratic mean* o "q.m") a μ. Esto se representa como:

$$x_n \xrightarrow{q.m} \mu$$

Además, se tiene que plim $x_n=\mu$.

Este último resultado se basa en la desigualdad de **Chebychev**, la cual establece que si x_n es una variable aleatoria y c y ε son constantes, entonces $\text{Prob}(|x_n-c|>\varepsilon)\leq E(x_n-c)^2/\varepsilon^2$.

Si hacemos $c=\mu_n$, tenemos que $\text{Prob}(|x_n-\mu_n|\geq\varepsilon)\leq E(x_n-\mu_n)^2/\varepsilon^2=\sigma_n^2/\varepsilon^2$. Si tomamos límites en ambos lados de la desigualdad cuando $n\to\infty$, tenemos:

$$\lim_{n\to\infty} \text{Prob}(|x_n-\mu_n|\geq\varepsilon)\leq \lim_{n\to\infty} \sigma_n^2/\varepsilon^2$$

Lo cual implica que plim $x_n=\mu$, dado que $\lim_{n\to\infty} \mu_n = \mu$ y $\lim_{n\to\infty}\sigma_n^2=0$.

Importante: Convergencia en media cuadrática implica convergencia en probabilidad, pero **no** viceversa. Utilicemos el Ejemplo 1 para ilustrar este punto:

$$E(x_n) = n\frac{1}{n} + 0\left(1-\frac{1}{n}\right) = 1;\ \text{Var}(x_n) = (n-1)^2\frac{1}{n} + (0-1)^2\left(1-\frac{1}{n}\right) = n$$

Entonces $\lim_{n\to\infty} E(x_n)=1$; $\lim_{n\to\infty} \text{Var}(x_n)=\infty$. Sin embargo, plim $x_n=0$.

3.- Estimador Consistente

Se dice que un estimador $\hat{\theta}$ de un parámetro θ es consistente si y sólo si:

$$\text{plim}\ \hat{\theta}=\theta$$

Caso Particular: Consistencia de la Media Muestral

La media muestral $\bar{x}$ de cualquier población con media finita μ y varianza finita σ^2 es un estimador consistente de μ.

Veamos por qué:

Sabemos que $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, donde x_1, ..., x_n es una muestra de una población cuya distribución tiene media y varianza finitas μ y σ^2, respectivamente. Entonces:

$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu$$

$$Var(\bar{x}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$, asumiendo que las x's son

independientes e idénticamente distribuidas (i.i.d).

De lo anterior, $\lim_{n\to\infty} E(\bar{x})=\mu$ y $\lim_{n\to\infty} Var(\bar{x})=0$. Por lo tanto, $\bar{x}$ converge en media cuadrática a μ. Ello implica que $\text{plim}\,\bar{x}=\mu$ ♦

Corolario

Con muestreo aleatorio, para cualquier función g(x), si E(g(x)) y Var(g(x)) son constantes finitas, se tiene que:

$$\text{plim}\,\frac{1}{n}\sum_{i=1}^{n} g(x_i) = E(g(x_i))$$

Demostración

Sea $y_i=g(x_i)$. El resultado se sigue de la consistencia de la media muestral.

Ejemplo 2

Sea $x \sim N(\mu, \sigma^2)$. Entonces $E(e^x)= e^{\mu+0.5\sigma^2}$ y $Var(e^x)= e^{2\mu+0.5\sigma^2} - e^{2\mu+\sigma^2}$. En base al corolario anterior, $\text{plim}\,\frac{1}{n}\sum_{i=1}^{n} e^{x_i} = e^{\mu+0.5\sigma^2}$ ♦

4.- Teorema de Slutsky

Para una función continua $g(x_n)$ que no es una función de n se tiene:

$\text{plim } g(x_n)=g(\text{plim } x_n)$.

5.- Reglas de la Probabilidad Límite

Estas son simplemente aplicaciones del teorema de Slutsky.

a) Escalares

Si x_n e y_n son variables aleatorias con $\text{plim } x_n=c$ y $\text{plim } y_n=d$, entonces:

$\text{plim}(x_n + y_n)=c + d$	regla de la suma
$\text{plim}(x_n\, y_n)=c\, d$	regla del producto
$\text{plim}\left(\frac{x_n}{y_n}\right)=\frac{c}{d}$	regla de la división (con $d\neq 0$).

Ejemplo 3

Supongamos que la media y varianza muestral del conjunto de variables aleatorias i.i.d x_1, ..,x_n con esperanza y varianza poblacional μ y σ^2, respectivamente, son estimadores consistentes. Esto es,

$$\text{plim } \bar{x} = \text{plim } \frac{1}{n}\sum_{i=1}^{n} x_i = \mu\,,\ \text{plim } s^2 = \text{plim } \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sigma^2.$$

Entonces, por el teorema de Slutsky:

$$\text{plim}\left(\frac{\bar{x}^2}{s^2}\right) = \frac{\text{plim}(\bar{x}^2)}{\text{plim}(s^2)} = \frac{(\text{plim}(\bar{x}))^2}{\text{plim}(s^2)} = \frac{\mu^2}{\sigma^2}\ \blacklozenge$$

b) Matrices

Sea $\mathbf{W}_n$ una matriz cuyos elementos son variables aleatorias, tal que plim$\mathbf{W}_n$=$\mathbf{\Omega}$, con $\mathbf{\Omega}$ matriz invertible. Entonces:

plim $\mathbf{W}_n^{-1}=\mathbf{\Omega}^{-1}$ regla de la matriz inversa

Si $\mathbf{X}_n$ e $\mathbf{Y}_n$ son matrices de variables aleatorias, tal que plim $\mathbf{X}_n$=$\mathbf{A}$ y plim $\mathbf{Y}_n$=$\mathbf{B}$, entonces

plim $(\mathbf{X}_n \mathbf{Y}_n)=\mathbf{A}\,\mathbf{B}$ regla de la matriz producto

6.- La Desigualdad de Jensen

Si $g(x_n)$ es una función cóncava de x_n, entonces:

$g(E(x_n)) \geq E(g(x_n))$

En tanto, si la función g(.) es convexa, entonces $g(E(x_n)) \leq E(g(x_n))$. Si la función g(.) es lineal, $g(E(x_n)) = E(g(x_n))$.

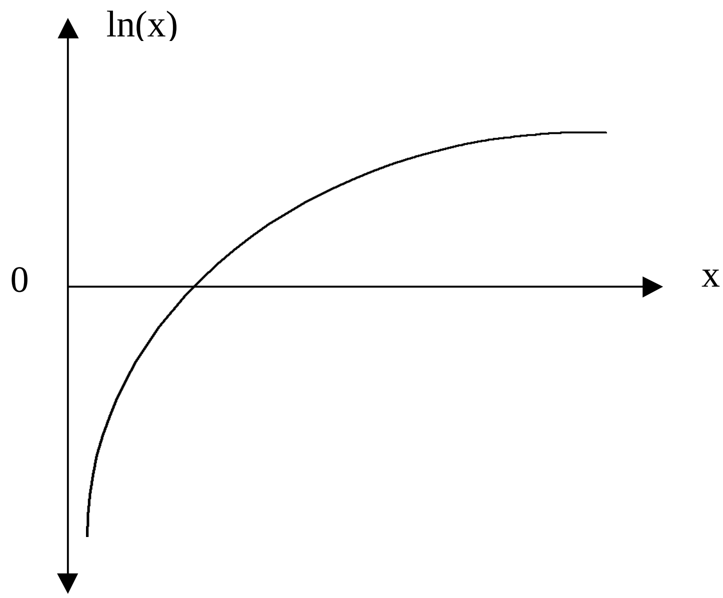
Veamos un ejemplo.

Ejemplo 4

Supongamos la siguiente función cóncava:

$g(x)=\ln(x)$, $x>0$

$$g'(x) = \frac{1}{x} > 0\,,\; g''(x) = -\frac{1}{x^2} < 0$$



Supongamos que x es una variable aleatoria discreta, tal que:

$$f(x) = \begin{cases} \frac{1}{2} & \text{si } x = 1 \\ \frac{1}{2} & \text{si } x = \frac{1}{2} \\ 0 & \text{si no} \end{cases}$$

$$E(x) = 1 \text{ x } \frac{1}{2} + \frac{1}{2} \text{ x } \frac{1}{2} = \frac{3}{4}.$$

Sea y=ln(x) una nueva variable aleatoria. Entonces:

$$f(y) = \begin{cases} \frac{1}{2} & \text{si } y = 0 \\ \frac{1}{2} & \text{si } y = -\ln(2) \\ 0 & \text{si no} \end{cases}$$

$$E(y) = 0 \text{ x } \frac{1}{2} - \ln(2)\text{x } \frac{1}{2} = -\ln\sqrt{2} \approx -0.35$$

De lo anterior es fácil verificar que ln E(x)≈-0.29>E(ln(x))≈–0.35♦

7.- Convergencia en Distribución

x_n converge en distribución a una variable aleatoria x con función distribución acumulada (f.d.a) F(x) si:

$\lim_{n\to\infty} (|F(x_n) - F(x)|) = 0$

en todos aquellos puntos de continuidad de F(x).

Esto se simboliza como $x_n \xrightarrow{d} x$.

Reglas para la Distribución Límite

1.- Si $x_n \xrightarrow{d} x$ y plim y_n=c, entonces:

$$x_n y_n \xrightarrow{d} cx$$

Además,

$$x_n + y_n \xrightarrow{d} c + x$$

$$\frac{x_n}{y_n} \xrightarrow{d} \frac{x}{c} \qquad \text{con } c \neq 0$$

2.- Si $x_n \xrightarrow{d} x$ y g(x_n) es una función continua, entonces $g(x_n) \xrightarrow{d} g(x)$.

3.-Si plim(x_n–y_n)=0, entonces x_n e y_n tienen la misma distribución límite.

Ejemplo 5

Supongamos una muestra de n observaciones i.i.d extraídas de la distribución x~N(0, σ^2). Sabemos de los cursos de inferencia estadística que:

$$\frac{\bar{x}}{s/\sqrt{n}} \sim t_{n-1}$$

donde $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Bajo ciertas condiciones de regularidad, se tiene que plim $s^2=\sigma^2$ y $\sqrt{n}\bar{x} \xrightarrow{d} N(0,\sigma^2)$. Entonces

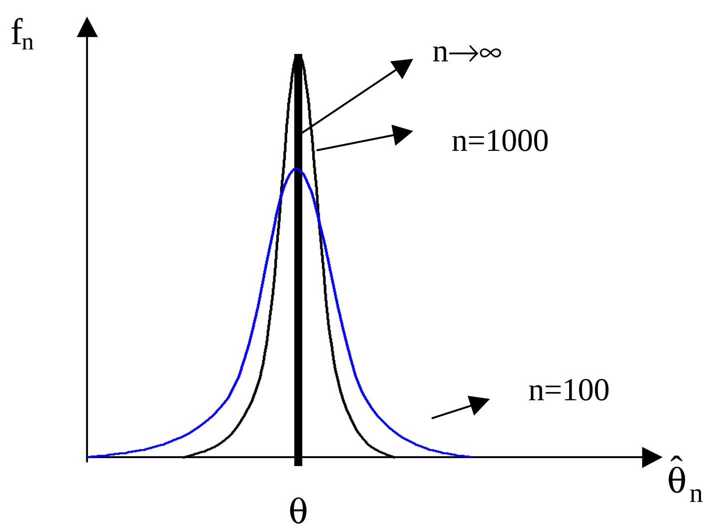
$$\frac{1}{s}\sqrt{n}\bar{x} \xrightarrow{d} N(0,1) \ \blacklozenge$$

Importante

Convergencia en probabilidad implica convergencia en distribución, pero no viceversa. Es decir, el concepto de convergencia en probabilidad es más fuerte. Veamos por qué. En primer término, si $\text{plim}(\hat{\theta}_n) = \theta$, entonces $\hat{\theta}_n \xrightarrow{d} \theta$. Ello, simplemente porque:

$$\lim_{n\to\infty} f(\hat{\theta}_n) = \begin{cases} 1 & \text{si } \hat{\theta}_n = \theta \\ 0 & \text{si no} \end{cases}$$

Gráficamente



(Para una demostración formal, ver Amemiya (1985) o White (1984)).

Por otra parte, convergencia en distribución no implica convergencia en probabilidad a una constante. Para probar tal aseveración, basta con dar un contraejemplo.

Supongamos que:

$$f(x_n) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{n+1} & \text{si } x_n = 1 \\ \dfrac{1}{2} - \dfrac{1}{n+1} & \text{si } x_n = 2 \end{cases}$$

Se tiene que $x_n \xrightarrow{d} x$, donde $f(x) = \begin{cases} \dfrac{1}{2} & \text{si } x = 1 \\ \dfrac{1}{2} & \text{si } x = 2 \end{cases}$

Es decir, x_n converge a una variable aleatoria pero no a una constante♦

Las distintas formas de convergencia se pueden resumir en el siguiente diagrama:

$$\begin{tikzcd} & \text{a. s} \arrow[d] & \\ \text{q. m} \arrow[r] & \text{p} \arrow[r] & \text{d} \end{tikzcd}$$

<u>Fuente</u>: Amemiya (1985)

Como veremos en las próximas secciones, una manera más informativa de describir un estimador cualquiera, $\hat{\theta}$, es a través de una **transformación estabilizante**:

$$z_n \equiv \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} f(z)$$

donde f(z) es una función distribución de probabilidades bien definida, con media y varianza finitas.

8.- <u>Teorema del Límite Central Univariado (Lindberg-Levy)</u>

Si x_1, x_2, ..., x_n constituye una muestra aleatoria de una distribución de probabilidad con una media finita μ y varianza σ^2 y $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, entonces:

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

9.- <u>Teorema del Límite Central con Varianzas Distintas (Lindberg-Feller)</u>

Sea $\{x_i\}_{i=1, \ldots, n}$ un conjunto de n variables aleatorias con medias y varianzas finitas μ_i y σ_i^2, respectivamente. Defínase $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2$.

Si $\lim_{n\to\infty} \frac{\max(\sigma_i)}{n\bar{\sigma}_n} = 0$ y $\lim_{n\to\infty} \bar{\sigma}_n^2 = \bar{\sigma}^2$, entonces:

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

donde $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$.

Estos dos teoremas también se pueden expresar en términos matriciales (ver Greene, capítulo 4).

10.- Distribución Asintótica de una Función de una Variable Aleatoria

Supongamos que $\sqrt{n}(z_n - \mu) \xrightarrow{d} N(0, \sigma^2)$. Entonces si $g(z_n)$ es una función continua que no depende de n, se tiene que:

$$\sqrt{n}(g(z_n) - g(\mu)) \xrightarrow{d} N(0, (g'(\mu))^2 \sigma^2)$$

Un bosquejo de la demostración es el siguiente. Por una expansión de Taylor de primer orden se tiene que:

$$g(z_n) \approx g(\mu) + g'(\mu)(z_n - \mu)$$

De ello, $\sqrt{n}(g(z_n) - g(\mu)) \approx g'(\mu)\sqrt{n}(z_n - \mu) \xrightarrow{d} N(0, (g'(\mu))^2 \sigma^2)$ ♦

Para analizar el caso **multivariado**, consideremos un vector $\mathbf{z}_n$ de variables aleatorias k x 1, tal que $\sqrt{n}(\mathbf{z}_n - \mu) \xrightarrow{d} N(\mathbf{0}, \Sigma)$. Si $\mathbf{g}(\mathbf{z}_n)$ es un vector de J funciones continuas de $\mathbf{z}_n$ que no dependen de n, entonces:

$$\sqrt{n}(\mathbf{g}(\mathbf{z}_n) - \mathbf{g}(\mu)) \xrightarrow{d} N(\mathbf{0}, \mathbf{C}\,\Sigma\,\mathbf{C}'),$$

donde **C** es una matriz j x k cuya j-ésima fila es el vector de derivadas parciales de la j-ésima función con respecto a $\mathbf{z}_n$, evaluado en μ:

$$\mathbf{C} = \begin{pmatrix} \dfrac{\partial g_1(\mu)}{\partial \mathbf{z}_n'} \\ \dfrac{\partial g_2(\mu)}{\partial \mathbf{z}_n'} \\ \dots \\ \dfrac{\partial g_n(\mu)}{\partial \mathbf{z}_{n'}} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial g_1(\mu)}{\partial z_{1n}} & \dfrac{\partial g_1(\mu)}{\partial z_{2n}} & \dots & \dfrac{\partial g_1(\mu)}{\partial z_{kn}} \\ \dfrac{\partial g_2(\mu)}{\partial z_{1n}} & \dfrac{\partial g_2(\mu)}{\partial z_{2n}} & \dots & \dfrac{\partial g_2(\mu)}{\partial z_{kn}} \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial g_J(\mu)}{\partial z_{1n}} & \dfrac{\partial g_J(\mu)}{\partial z_{2n}} & \dots & \dfrac{\partial g_J(\mu)}{\partial z_{kn}} \end{pmatrix},$$

$$\mathbf{z}_n = \begin{pmatrix} z_{1n} \\ z_{2n} \\ \dots \\ z_{kn} \end{pmatrix}$$

Aplicaciones: Consistencia y Normalidad Asintótica de Mínimos Cuadrados Ordinarios (MICO)

Consideremos el modelo lineal clásico expresado en términos matriciales:

$\mathbf{y}=\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ $\qquad$ $E(\boldsymbol{\varepsilon}|\mathbf{X})=0$, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')=\sigma^2$ I, con σ^2 constante finita.

Asumamos que $\lim_{n\to\infty} \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q}$, matriz positiva definida e invertible, donde, por simplicidad, se asume que $\mathbf{X}$ es una matriz de variables no estocásticas.

El estimador MICO viene dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \boldsymbol{\beta} + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon})$$

Entonces:

$$\text{plim}\,\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \lim_{n\to\infty}(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}\,\text{plim}\,(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}) = \boldsymbol{\beta} + \mathbf{Q}^{-1}\,\text{plim}\,(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}),$$

por las propiedades de probabilidad límite descritas en secciones anteriores.

Pero $\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i = \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i \equiv \overline{\mathbf{w}}$,

donde $\mathbf{x}_i$ es el vector 1 x k correspondiente a la i-ava fila de la matriz $\mathbf{X}$ y $\mathbf{w}_i\equiv\mathbf{x}_i\varepsilon_i$.

Se tiene que:

$$E(\overline{\mathbf{w}}) = \frac{1}{n}\sum_{i=1}^{n}E(\mathbf{w}_i) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i E(\varepsilon_i) = \mathbf{0}$$

$$\text{Var}(\overline{\mathbf{w}}) = E(\overline{\mathbf{w}}\,\overline{\mathbf{w}}') = \frac{1}{n^2}\mathbf{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X} = \frac{\sigma^2}{n}\frac{\mathbf{X}'\mathbf{X}}{n}.$$

De ello, $\lim_{n\to\infty} \mathrm{Var}(\overline{\mathbf{w}}) = \lim_{n\to\infty}\left(\frac{\sigma^2}{n}\right)\lim_{n\to\infty}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right) = 0 \text{ x } \mathbf{Q}^{-1} = \mathbf{0}$.

Esto implica que $\overline{\mathbf{w}}$ converge en media cuadrática a **0** y, por lo tanto, plim $\overline{\mathbf{w}}$ =**0**. Es decir, plim $\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{0}$.

En consecuencia,

$$\text{plim}\,\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{Q}^{-1}\,\mathbf{0} = \boldsymbol{\beta}.$$

Esto es, el estimador MICO es **consistente**.

El supuesto de **X** no estocásticos se puede relajar fácilmente. En verdad, se reemplaza el supuesto $\lim_{n\to\infty}\frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q}$ por $\text{plim}\,\frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q}$ y se procede de manera análoga.

Para establecer **normalidad asintótica**, notemos que la igualdad $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon})$ puede ser reescrita como:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon}).$$

Dado que $\lim_{n\to\infty}\frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q}$, necesitamos establecer la distribución límite de $\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon} = \sqrt{n}\,\overline{\mathbf{w}}$. Para ello, hacemos uso del teorema del límite central de Lindberg-Feller.

Notemos que $\overline{\mathbf{w}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i$ es el promedio de n variables aleatorias independientes $\mathbf{x}_i\varepsilon_i$ con media **0** y varianza Var($\mathbf{x}_i\varepsilon_i$)= $\mathbf{x}_i$ Var(ε_i) $\mathbf{x}_i'$=σ^2 $\mathbf{x}_i$ $\mathbf{x}_i'$.

Por lo tanto,

$$\mathrm{Var}(\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon}) = \mathrm{Var}(\sqrt{n}\overline{\mathbf{w}}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(\varepsilon_i \mathbf{x}_i) = \frac{\sigma^2}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \sigma^2\frac{\mathbf{X}'\mathbf{X}}{n}$$

$$\rightarrow \sigma^2\mathbf{Q}$$

a medida que $n\rightarrow\infty$.

Entonces $\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon} = \sqrt{n}\overline{\mathbf{w}} \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q})$.

Por propiedades de distribuciones límites:

$$(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon}) \xrightarrow{d} N(\mathbf{Q}^{-1}\mathbf{0}, \mathbf{Q}^{-1}(\sigma^2\mathbf{Q})\mathbf{Q}^{-1}) = N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$$

Entonces

$$\sqrt{n}(\hat{\beta}_{MICO} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$$

Para una discusión de la derivación con X estocásticos, ver Greene, capítulo 6.

Nota: Consistencia y Normalidad Asintótica bajo Mínimos Cuadrados Generalizados (MCG)

Recordemos que $\hat{\beta}_{MCG} = (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{y}_* = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$, con $\mathbf{X}_*=\mathbf{PX}$, $\mathbf{y}_*=\mathbf{Py}$, $\mathbf{P}'=\mathbf{C}\boldsymbol{\Lambda}^{-1/2}$, $\boldsymbol{\Omega}=\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')=\sigma^2\boldsymbol{\Omega}$, $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$.

1) El estimador de mínimos cuadrados generalizado es consistente si:

$\mathrm{plim}(\frac{\mathbf{X}_*'\mathbf{X}_*}{n}) = \mathbf{Q}_*$ es una matriz positiva definida e invertible.

Es decir, lo que necesitamos es que los datos transformados sean bien comportados.

2) Se tiene $\hat{\beta}_{MCG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$

Entonces

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) = \left(\frac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{n}\right)^{-1} \frac{\mathbf{X}'\Omega^{-1}\boldsymbol{\varepsilon}}{\sqrt{n}}.$$

Si $\left(\frac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{n}\right)^{-1} \xrightarrow{p} \mathbf{Q}_*^{-1}$ y $\frac{\mathbf{X}'\Omega^{-1}\boldsymbol{\varepsilon}}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q}_*^{-1})$, entonces:

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q}_*^{-1}).$$

<u>Resultados Utiles</u>

1.- <u>Consistencia de</u> $\hat{\sigma}^2$. Consideremos nuevamente el modelo lineal clásico, en el cual $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, $E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}$. Recordemos que $\hat{\sigma}^2 = \frac{1}{n-k}\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$, donde $\boldsymbol{\varepsilon}=\mathbf{y}-\mathbf{X}\boldsymbol{\beta}$, error poblacional, $\mathbf{M}=I-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$, $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, error estimado.

(¿Por qué $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$? Notemos que $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{M}\mathbf{y}$, entonces $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{y}'\mathbf{M}\mathbf{y}$ porque M es idempotente. Pero $\mathbf{y}'\mathbf{M}\mathbf{y}=(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})'\mathbf{M}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}'\mathbf{M}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}\boldsymbol{\varepsilon}+\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$, porque $\mathbf{M}\mathbf{X}=\mathbf{0}$).

$$\hat{\sigma}^2 = \frac{1}{n-k}\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} = \frac{1}{n-k}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon})$$

$$= \frac{n}{n-k}\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n} - \left(\frac{\boldsymbol{\varepsilon}'\mathbf{X}}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n}\right)\right)$$

Tenemos que $\text{plim}\left(\frac{\boldsymbol{\varepsilon}'\mathbf{X}}{n}\right) = \mathbf{0}$, $\text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = \mathbf{Q}^{-1}$, y $\lim_{n\to\infty} \frac{n}{n-k} = 1$ para un k fijo. Por ello,

$$\text{plim}\,\hat{\sigma}^2 = \text{plim}\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n} = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$$

Supongamos que los ε′s son i.i.d. tal que $E(\varepsilon_i^2)=\sigma^2$ y $E(\varepsilon_i^4)=\phi<\infty$. Entonces $Var(\varepsilon_i^2)= E(\varepsilon_i^4)-E^2(\varepsilon_i^2)=\phi-\sigma^4<\infty$. De esto se deduce que:

$$\text{plim}\ \hat{\sigma}^2 = \text{plim}\ \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = E(\varepsilon_i^2) = \sigma^2$$

Por lo tanto, $\hat{\sigma}^2$ es un estimador consistente de σ^2 ♦

2.- Distribución Asintótica del Estadígrafo t

$$t_k = \frac{\hat{\beta}_k - \beta_k}{(\hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}_{kk})^{1/2}}$$

donde $(\mathbf{X'X})_{kk}^{-1}$ es el k-avo elemento de la diagonal de $(\mathbf{X'X})^{-1}$.

Dado que $\text{plim}\,\hat{\sigma}^2 = \sigma^2$ y $\sqrt{n}(\hat{\beta}_k - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}_{kk})$, donde $\mathbf{Q}_{kk}^{-1}$ es el k-avo elemento de la diagonal de $\mathbf{Q}^{-1}$, el estadígrafo $t_k \xrightarrow{d} N(0,1)$ ♦

3.- Distribución Asintótica del Test de Restricciones Lineales

Supongamos que queremos contrastar un conjunto de J restricciones lineales:

H_0: $\mathbf{R}\boldsymbol{\beta}=\mathbf{q}$ H_1: $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{q}$,

donde **R** es una matriz J x k, **β** es un vector k x 1 y **q** es un vector J x 1.

Por ejemplo,

1) Un subconjunto de los coeficientes es igual a cero:

H_0: $\beta_1=0$, $\beta_2=0$, $\beta_3=0$.

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}_{3\,x\,k}, \quad \mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3\,x\,1}.$$

2) Varias restricciones al mismo tiempo: H_0: $\beta_2+\beta_3=1$, $\beta_4+\beta_6=0$, $\beta_5+\beta_6=0$.

$$\mathbf{R}=\begin{bmatrix}0&1&1&0&0&0&0&\ldots&0\\0&0&0&1&0&1&0&\ldots&0\\0&0&0&0&1&1&0&\ldots&0\end{bmatrix}_{3\text{ x }k}\quad,\ \mathbf{q}=\begin{bmatrix}1\\0\\0\end{bmatrix}_{3\text{ x }1}.$$

Sabemos de cursos anteriores que dicho conjunto de J restricciones puede ser contrastado con el siguiente estadígrafo:

$$F=\frac{(\mathbf{R}\hat{\beta}-\mathbf{q})'(\mathbf{R}\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\beta}-\mathbf{q})}{J}$$

donde $\hat{\beta}$ es el estimador MICO no restringido.

Este se distribuye F(J, n-k) bajo normalidad de los errores poblacionales del modelo lineal. No obstante, aun cuando el supuesto de normalidad no se satisfaga, es posible obtener la distribución asintótica del estadígrafo. Específicamente, en muestras grandes se tiene que:

$$J\text{ x }F=(\mathbf{R}\hat{\beta}-\mathbf{q})'(\mathbf{R}\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\beta}-\mathbf{q})\xrightarrow{d}\chi^2(J)$$

<u>Demostración</u>

Sabemos que $\sqrt{n}(\hat{\beta}-\hat{\beta})\xrightarrow{d}N(\mathbf{0},\sigma^2\mathbf{Q}^{-1})$. Entonces, dado que **R** es una matriz de constantes,

$$\sqrt{n}\mathbf{R}(\hat{\beta}-\beta)\xrightarrow{d}N(\mathbf{0},\mathbf{R}(\sigma^2\mathbf{Q}^{-1})\mathbf{R}')$$

Pero $\mathbf{R}\beta=\mathbf{q}$ bajo H_0. Entonces

$$\sqrt{n}(\mathbf{R}\hat{\beta}-\mathbf{q})\xrightarrow{d}N(\mathbf{0},\sigma^2\mathbf{R}\mathbf{Q}^{-1}\mathbf{R}')$$

Sea $\mathbf{Z}_{Jx1}=\sqrt{n}(\mathbf{R}\hat{\beta}-\mathbf{q})$ y $\mathbf{P}_{JxJ}=\sigma^2\mathbf{R}\mathbf{Q}^{-1}\mathbf{R}'$, con **P** matriz positiva definida. Sabemos que **P** se puede descomponer en $\mathbf{P}=\mathbf{T}\Lambda\mathbf{T}'$, donde **T** es la matriz de vectores propios de **P** y **Λ** es una matriz diagonal con los valores propios de **P** asociados a cada vector propio. Por lo tanto, $\mathbf{P}^{-1/2}=\mathbf{T}\mathbf{\Lambda}^{-1/2}\mathbf{T}'$.

Si $\mathbf{z} \xrightarrow{d} N(\mathbf{0},\mathbf{P})$, entonces

$$\mathbf{P}^{-1/2}\mathbf{z} \xrightarrow{d} N(\mathbf{0},\mathbf{P}^{-1/2}\mathbf{P}(\mathbf{P}^{-1/2})') = N(\mathbf{0},\mathbf{I}_{JxJ})$$

Sabemos, además, que la suma de J variables i.i.d. normales estándar al cuadrado se distribuye chi-cuadrado con J grados de libertad. (Cada variable normal estándar al cuadrado se distribuye chi-cuadrado con un grado de libertad).

Por lo tanto,

$$(\mathbf{P}^{-1/2}\mathbf{z})'(\mathbf{P}^{-1/2}\mathbf{z}) = \mathbf{z}'\mathbf{P}^{-1}\mathbf{z} \xrightarrow{d} \chi^2(J)$$

Ahora, dado que $\text{plim}(\frac{\mathbf{X}'\mathbf{X}}{n})^{-1} = \mathbf{Q}^{-1}$ y $\text{plim}\,\hat{\sigma}^2 = \sigma^2$, tenemos que $J \text{ x } F = \mathbf{z}'\mathbf{P}^{-1}\mathbf{z} \xrightarrow{d} \chi^2(J)$ ♦

Referencias

Amemiya T.(1985), *Advanced Econometrics*. Harvard University Press
Davidson, J. y J. MacKinnon (1993), *Estimation and Inference in Econometrics*. New York: Oxford University Press.
Greene W. (1997), *Econometric Analysis*. Prentice Hall, tercera edición.
White, H. (1984), *Asymptotic Theory for Econometricians*. Academic Press