Fundamentos de Econometría

Ignacio Lobato

ITAM

Teoría Asintótica

- En la sección anterior se definió a la media muestral y se vio que la distribución de la misma en una muestra aleatoria depende de la población de origen y el tamaño muestral.
- En particular, se vio que para cualquier población, $E(\overline{X}) = \mu$ y $V(\overline{X}) = \sigma^2/n$. Note que la esperanza de $\overline{X}$ se mantiene constante independientemente del tamaño muestral pero su varianza va decreciendo cuando la muestra se hace más grande.
- En esta sección se desarrollará información adicional respecto a la distribución de $\overline{X}$, que será válida para cualquier población.

Teoría Asintótica

- En esta sección se estudiarán los siguientes 3 resultados fundamentales:
- **1. Ley de Grandes Números:** Dada una media muestral denotada por $\overline{X}$, el *límite de probabilidad* de $\overline{X}$ es μ, donde μ denota la media poblacional asociada a $\overline{X}$.
- **2. Teorema del Límite Central:** La *distribución límite* de $Z = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$ es $N(0,1)$.
- **3.** La *distribución asintótica* de $\overline{X}$ es $N(\mu, \sigma^2/n)$

Secuencias de Estadísticos Muestrales

- Sea $\overline{X}_n$ una media muestral indexada por su tamaño muestral. Consideremos una secuencia de estos estadísticos muestrales, tal que cada uno de ellos tiene su propia distribución y momentos.
- De forma más general, sea T_n una secuencia de variables aleatorias con distribuciones acumuladas $G_n(t) = Pr(T_n \leq t)$, esperanzas $E(T_n)$ y varianzas $V(T_n)$.
- Cuando se hable de límites en esta sección, se hará referencia a que el tamaño de la muestra se va haciendo cada vez más grande hasta tender al infinito.
- De este modo, definimos tres tipos de convergencia:

Convergencia en Probabilidad

- T_n converge en probabilidad a c (constante) si:

$$\lim_{n \to +\infty} G_n(t) = 0 \quad \text{para todo } t < c \text{ y}$$

$$\lim_{n \to +\infty} G_n(t) = 1 \quad \text{para todo } t \geq c$$

Denotamos esto por $T_n \xrightarrow{p} c$ o como $plim T_n = c$.

- Sea $A_n = \{\mid T_n - c \mid \geq \epsilon\}$ donde $\epsilon \geq 0$. Entonces:

$$Pr(A_n) = 1 - G_n(c + \epsilon) + Pr(T_n = c + \epsilon) + G_n(c - \epsilon)$$

Entonces $T_n \xrightarrow{p} c$ sí y solo sí $\lim_{n \to +\infty} Pr(A_n) = 1 - 1 + 0 + 0 = 0$.

- Por lo tanto, una forma equivalente de definir la convergencia en probabilidad de T_n a c es:

$$\lim_{n \to +\infty} Pr(\mid T_n - c \mid \geq \epsilon) = 0$$

Para todo $\epsilon > 0$

- Si existe una constante c, tal que $\lim_{n\to+\infty} E(T_n - c)^2 = 0$, decimos que T_n converge en media cuadrática a c, lo cual se denotará por $T_n \overset{m}{\to} c$. A partir de esta definición, se cumplen dos propiedades inmediatas:

  C1. Si T_n es una secuencia de variables aleatorias con $\lim_{n\to+\infty} E(T_n) = c$ y $\lim_{n\to+\infty} V(T_n) = 0$, entonces T_n converge en media cuadrática a c.

  Demostración: Podemos escribir $E(T_n - c)^2$ como:

$$E(T_n - c)^2 = E(T_n - E(T_n) + E(T_n) - c)^2 = V(T_n) + [E(T_n) - c]^2$$

  Tomando límites tendremos que $T_n \overset{m}{\to} c$.

- **C2.** Si $T_n \stackrel{m}{\rightarrow} c$, entonces $T_n \stackrel{p}{\rightarrow} c$.

  Demostración:

  Sea $A_n = \{\mid T_n - c \mid \geq \epsilon\}$, donde $\epsilon > 0$. Aplicando la desigualdad de Chebyshev:

  $$0 \leq Pr(A_n) \leq E(T_n - c)^2/\epsilon^2$$

  Luego tomando límites se tiene que:

  $$0 \leq \lim_{n \to +\infty} Pr(A_n) \leq 0$$

  Por lo tanto, $\lim_{n \to +\infty} Pr(A_n) = 0$, lo cual implica que $T_n \stackrel{p}{\rightarrow} c$.

Convergencia en Distribución

- Si existe alguna cdf G(t) (fija), tal que $\lim_{n\to+\infty} G_n(t) = G(t)$ para todo t en el cual G(.) es continua, entonces decimos que T_n converge en distribución a G(.), lo cual denotaremos por $T_n \xrightarrow{d} G(.)$.
- Note que la convergencia en probabilidad es un caso especial de convergencia en distribución en la cual la distribución límite es degenerada.

Teoría Asintotica - Media Muestral

- Los siguientes conceptos se aplican a una secuencia de medias muestrales provenientes de muestras aleatorias de cualquier población:

  Ley de los Grandes Números: En una muestra aleatoria de cualquier población con $E(X) = \mu$ y $V(X) = \sigma^2$, la media muestral converge en probabilidad a la media poblacional ($\overline{X} \xrightarrow{p} \mu$).

  Demostración: Tenemos que $E(\overline{X}_n) = \mu$ y $V(\overline{X}_n) = \sigma^2/n$, entonces:

$$\lim_{n \to +\infty} \overline{X}_n = \mu \quad \text{y} \quad \lim_{n \to +\infty} V(\overline{X}_n) = 0$$

  Por lo tanto, de **C1**:

$$\overline{X}_n \xrightarrow{m} \mu$$

  y por **C2**:

$$\overline{X}_n \xrightarrow{p} \mu$$

Teoría Asintótica - Media Muestral

- **Teorema Central del Límite (TCL):** En una muestra aleatoria de cualquier población con $E(X) = \mu$ y $V(X) = \sigma^2$, la media muestral estandarizada $Z = \sqrt{n}(\overline{X} - \mu)/\sigma \xrightarrow{d} N(0,1)$. Equivalentemente, $\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$.

- Este teorema es empleado para aproximar la función de distribución acumulada de una media muestral. Así, si la función de distribución acumulada de $\overline{X}_n$ es $F_n(.)$, se tiene que:

$$F_n(c) = Pr(\overline{X}_n \leq c) = Pr[\sqrt{n}(\overline{X}_n - \mu)/\sigma \leq \sqrt{n}(c - \mu)/\sigma]$$

$$= Pr(Z_n \leq c^*) \approx \Phi(c^*)$$

Teoría Asintótica - Momentos Muestrales

- Los resultados aplicables a la media muestral son válidos para toda una clase de estadísticos que pueden ser interpretados como medias muestrales en una muestra aleatoria. Así, por ejemplo, para cualquier $r \in \mathbb{N}$:

$$\frac{\sum_i X_i^r}{n} \xrightarrow{p} E(X^r)$$

$$\sqrt{n}\left(\frac{\sum_i X_i^r}{n} - E(X^r)\right) / \sqrt{(E(X^{2r}) - E^2(X^r))} \xrightarrow{d} N(0,1)$$

$$\frac{\sum_i X_i^r}{n} \xrightarrow{d} N\left[E(X^r), (E(X^{2r}) - E^2(X^r))\right]$$

- Lo mismo aplica a los momentos centrados en la media poblacional, ya que esta última es una constante.

Teoría Asintótica - Funciones de Momentos Muestrales

- Usualmente estamos interesados en estadísticos que no pueden ser interpretados directamente como medias muestrales.
- Un ejemplo de este tipo de estadístico es el caso de la varianza muestral S^2, que puede escribirse de la siguiente manera:

$$S^2 = \sum_i (X_i - \overline{X})^2/n = \sum_i (X_i - \mu)^2/n - (\overline{X} - \mu)^2$$

  La cual es una función de un momento muestral centrado en la media poblacional y de la media muestral.

Teoría Asintótica - Funciones de Momentos Muestrales

- Para derivar los momentos y distribuciones asintóticas de funciones de medias muestrales, se utilizan los *Teoremas de Slutsky*.
- Sean T_n, V_n y W_n secuencias de variables aleatorias, mientras que la función h(.) y la constante c no dependen de n. Los teoremas de Slutsky son los siguientes:

  S1: Si $T_n \xrightarrow{p} c$ y $h(T_n)$ es continua en c, entonces $h(T_n) \xrightarrow{p} h(c)$.

  S2: Si $V_n \xrightarrow{p} c_1$ y $W_n \xrightarrow{p} c_2$, y $h(V_n, W_n)$ es continua en (c_1, c_2), entonces $h(V_n, W_n) \xrightarrow{p} h(c_1, c_2)$.

  S3: Si $V_n \xrightarrow{p} c$ y W_n tiene una distribución límite, luego la distribución límite de $(V_n + W_n)$ es la misma que la correspondiente a $(c + W_n)$.

  S4: Si $V_n \xrightarrow{p} c$ y W_n tiene una distribución límite, entonces la distribución límite de $(V_n W_n)$ es la misma que la correspondiente a cW_n.

S5. Método Delta: Si $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \phi^2)$ y $U_n = h(T_n)$ es continuamente diferenciable en θ, entonces:

$$\sqrt{n}\,[U_n - h(\theta)] \xrightarrow{d} N\{0, [h'(\theta)]^2\phi^2\}$$

Este resultado es consecuencia de una aproximación de Taylor de la función U_n alrededor de θ. Así:

$$U_n = h(T_n) = h(\theta) + h'(\theta)(T_n - \theta)$$

utilizando la convergencia en distribución de U_n, y aplicando S4 obtenemos S5.

Algunos Ejemplos - Varianza Muestral

- Sabemos que la varanza muestral se puede escribir de la siguiente manera:

$$S^2 = \sum_i (X_i - \overline{X})^2/n = \sum_i (X_i - \mu)^2/n - (\overline{X} - \mu)^2 = h(\frac{\sum_i (X_i - \mu)^2}{n}, \overline{X})$$

Luego por la Ley de Grandes Números, $\frac{\sum_i (X_i - \mu)^2}{n} \xrightarrow{p} E(X - \mu)^2 = \sigma^2$ y $\overline{X} \xrightarrow{p} E(X) = \mu$. Luego por S2:

$$S^2 = h(\frac{\sum_i (X_i - \mu)^2}{n}, \overline{X}) \xrightarrow{p} h(\sigma^2, E(X)) = \sigma^2 - (\mu - \mu)^2 = \sigma^2$$

Por lo tanto, $S^2 \xrightarrow{p} \sigma^2$.

Algunos Ejemplos - Varianza Muestral

- Por otro lado, tenemos que:

$$\sqrt{n}(S^2-\sigma^2)=\sqrt{n}(\frac{\sum_i (x_i-\mu)^2}{n}-\sigma^2)-U^2$$

con $U^2=\sqrt{n}(\overline{X}-\mu)$. Además note que:

$$E(U)=0 \quad \text{y} \quad V(U)=\sqrt{n}V(\overline{X})=\sigma^2/\sqrt{n}$$

Y note que $\lim_{n\to+\infty} E(U)=0$ y $\lim_{n\to+\infty} V(U)=0$. Por lo tanto, $U\xrightarrow{m}0$, lo cual implica que $U\xrightarrow{p}0$. Luego por S1, $U^2\xrightarrow{p}0$. Finalmente por S3, podemos concluir que:

$$\sqrt{n}(S^2-\sigma^2)\xrightarrow{d}N(0,E(X-\mu)^4-\sigma^4)$$

Algunos Ejemplos - Ratio-t Muestral

- Sean:

$$U = \sqrt{n}(\overline{X} - \mu)/S \quad \text{y} \quad Z = \sqrt{n}(\overline{X} - \mu)/\sigma$$

Entonces $U = (\sigma/S)Z$.

Además, sabemos por el TCL que:

$$Z \xrightarrow{d} N(0,1)$$

Asimismo, note que:

$$\sigma/S \xrightarrow{p} 1$$

Pues $S^2 \xrightarrow{p} \sigma^2$ por S1.

Finalmente, por S4:

$$U \xrightarrow{d} N(0,1)$$

Teoría Asintótica - El Caso Bivariado

- Ahora estudiamos la teoría asintótica en un contexto bivariado.
- Se entenderá que para un vector aleatorio, convergencia en probabilidad significa que cada uno de sus componentes converge en probabilidad.
- En el mismo contexto, convergencia en distribución significa que la secuencia de funciones de distribución acumulada tiene un límite en alguna función de distribución acumulada conjunta.
- De este modo, los teoremas básicos en el contexto bivariado son: **1) La ley bivariada de grandes números y 2) El teorema central del límite bivariado.**

Teoría Asintótica - El Caso Bivariado

- Sean $\overline{X}$ y $\overline{Y}$ un par de medias muestrales. Entonces se cumple que:

  1) La ley bivariada de grandes números (LBGN): En una muestra aleatoria de cualquier población bivariada, el vector de medias muestrales satisface:

$$(\overline{X}, \overline{Y}) \xrightarrow{p} (\mu_x, \mu_y)$$

  2) El teorema central del límite bivariado (TCLB): En una muestra aleatoria de cualquier población bivariada, el vector de medias muestrales estandarizadas satisface:

$$\left[\sqrt{n}(\overline{X} - \mu_x)/\sigma_x, \sqrt{n}(\overline{Y} - \mu_y)/\sigma_y\right] \xrightarrow{d} SBVN(\rho)$$

  con $\rho = \sigma_{xy}/(\sigma_x \sigma_y)$. Equivalentemente, en una muestra aleatoria de cualquier población bivariada, el par de medias muestrales satisface:

$$(\overline{X}, \overline{Y}) \xrightarrow{d} BVN(\mu_x, \mu_y, \sigma_x^2/n, \sigma_y^2/n, \sigma_{xy}/n)$$

Teoría Asintótica - El Caso Bivariado

- La LGNB y el TCLB se pueden aplicar para cualquier par de momentos que puedan ser interpretados como un par de medias muestrales en una muestra aleatoria.
- Para el caso de funciones de medias muestrales, los teoremas de Slutsky (S1-S4) se pueden extender de una manera trivial al caso bivariado.

- En el caso del Método Delta, este también puede ser generalizado de la siguiente manera:

  Método Delta Bivariado: Si $(T_1, T_2) \xrightarrow{d} BVN(\theta_1, \theta_2, \phi_1^2/n, \phi_2^2/n, \phi_{12}/n)$ y $U = h(T_1, T_2)$ es continuamente diferenciable en el punto (θ_1, θ_2), entonces:

  $$U \xrightarrow{d} N(h(\theta_1, \theta_2), \phi^2/n)$$

  donde,

  $$\phi^2 = h_1^2\phi_1^2 + h_2^2\phi_2^2 + 2h_1h_2\phi_{12}$$

  $$h_1 = h_1(\theta_1, \theta_2) = \partial h(T_1, T_2)/\partial T_1$$

  $$h_2 = h_2(\theta_1, \theta_2) = \partial h(T_1, T_2)/\partial T_2$$

  En otras palabras, la distribución asintótica de U es obtenida por su aproximación de Taylor alrededor del punto (θ_1, θ_2).

Algunos Ejemplos - Ratio de Medias Muestrales

- Sea $T = \overline{X}/\overline{Y}$ y sea $\mu_y \neq 0$. Derivamos las propiedades asintóticas de T.

$$T = \overline{X}/\overline{Y} = h(\overline{X}, \overline{Y})$$

Pero sabemos que $\overline{X} \xrightarrow{p} \mu_x$ y $\overline{Y} \xrightarrow{p} \mu_y$. Sea $\theta = \mu_x/\mu_y$. Entonces:

$$T \xrightarrow{p} \theta$$

Ahora, aplicando una aproximación de Taylor a T alrededor de (μ_x, μ_y), tenemos que:

$$T \approx \theta + (1/\mu_y)(\overline{X} - \mu_x) - (\theta/\mu_y)(\overline{Y} - \mu_y)$$

Entonces la varianza asintótica de T será ϕ^2/n, donde:

$$\phi^2 = (1/\mu_y)^2(\sigma_x^2 + \theta^2\sigma_y^2 - 2\theta\sigma_{xy})$$

Luego:

$$T \xrightarrow{d} N(\theta, \phi^2/n)$$

Algunos Ejemplos - Covarianza Muestral

- Sabemos que la covarianza muestral es la siguiente:

$$S_{xy} = (1/n)\sum_i (X_i - \overline{X})(Y_i - \overline{Y}) = M_{11}$$

Para encontrar las propiedades asintóticas de S_{xy}, primero definimos la covarianza muestral 'ideal':

$$M_{11}^* = (1/n)\sum_i (X_i - \mu_x)(Y_i - \mu_y) = (1/n)\sum_i V_i = \overline{V}$$

Y note que:

$$E(M_{11}^*) = \sigma_{xy} \quad \text{y}$$

$$V(M_{11}^*) = \left(E((X - \mu_x)^2(Y - \mu_y)^2) - \sigma_{xy}^2\right)/n = (\mu_{22} - \mu_{11}^2)/n$$

Entonces

$$\mu_{11}^* \xrightarrow{p} \sigma_{xy} = \mu_{11} \quad \text{y} \sqrt{n}(M_{11}^* - \mu_{11}) \xrightarrow{d} N(\mu_{11}, \mu_{22} - \mu_{11}^2)$$

Algunos Ejemplos - Covarianza Muestral

Volviendo a la covarianza muestral note que:

$$S_{xy} = M_{11} = M_{11}^* - (\overline{X} - \mu_x)(\overline{Y} - \mu_y)$$

y note que

$$E(S_{xy}) = \sigma_{xy} - C(\overline{X}, \overline{Y}) = (1 - 1/n)\sigma_{xy}$$

$$V(S_{xy}) = (n-1)^2(\mu_{22} - \mu_{11}^2)/n^3 + 2(n-1)(\sigma_x^2\sigma_y^2)/n^3$$

Por otro lado, tenemos que:

$$M_{11} \xrightarrow{p} \sigma_{xy} \quad , \quad (\overline{X} - \mu_x) \xrightarrow{p} 0 \quad \text{y} \quad (\overline{Y} - \mu_y) \xrightarrow{p} 0$$

Entonces tendremos que por Slutsky:

$$S_{xy} \xrightarrow{p} \sigma_{xy} \quad \text{y} \quad \sqrt{n}(S_{xy} - \sigma_{xy}) \xrightarrow{d} N(0, \mu_{22} - \mu_{11}^2)$$

$$S_{xy} \xrightarrow{d} N(\sigma_{xy}, (\mu_{22} - \mu_{11}^2)/n)$$

Algunos Ejemplos - Mejor Predictor Lineal (BLP)

- En secciones anteriores se vio la proyección lineal poblacional de Y sobre X en una distribución bivariada, denotada por $E^*(Y/X) = \alpha + \beta X$, con:

$$\beta = \sigma_{xy}/\sigma_x^2 \quad \text{y} \quad \alpha = \mu_y - \beta\mu_x$$

- Los 'análogos muestrales' de β y α serían los siguientes:

$$B = S_{xy}/S_x^2 = M_{11}/M_{20} \quad \text{y} \quad A = \overline{Y} - B\overline{X}$$

- Ahora definamos la pendiente 'ideal' en la muestra para el mejor predictor lineal de la siguiente forma:

$$B^* = M_{11}^*/M_{20}^* = \overline{V}/\overline{W}$$

donde,

$$M_{11}^* = (1/n)\sum_i X_i^* Y_i^* = \overline{V} \quad \text{y} \quad M_{20}^* = (1/n)\sum_i X_i^2 = \overline{W}$$

X_i^* y Y_i^* denotan a X_i y Y_i centradas en su media poblacional.

Algunos Ejemplos - Mejor Predictor Lineal (BLP)

Note que:

$$\mu_v = E(X^* Y^*) = \sigma_{xy} \quad \text{y} \quad \mu_w = E(X^{*2}) = \sigma_x^2$$

y que $\beta = \mu_v / \mu_w$. Entonces, tenemos que:

$$B^* \xrightarrow{p} \sigma_{xy}/\sigma_x^2 = \beta$$

Y aplicando el método delta bivariado, tendremos que:

$$B^* \xrightarrow{d} N(\beta, \phi^2/n)$$

con

$$\phi^2 = (1/\mu_w)^2(\sigma_v^2 + \beta^2\sigma_w^2 - 2\beta\sigma_{vw})$$

donde,

$$\sigma_v^2 = V(V) = E(V^2) - E^2(V) = E(X^{*2}Y^{*2}) - E^2(X^*Y^*) = \mu_{22} - \mu_{11}^2$$
$$\sigma_w^2 = V(W) = E(W^2) - E^2(W) = E(X^{*4}) - E^2(X^{*2}) = \mu_{40} - \mu_{20}^2$$
$$\sigma_{vw} = C(V, W) = E(X^{*3}Y^*) - E(X^*Y^*)E(X^{*2}) = \mu_{31} - \mu_{11}\mu_{20}$$

Algunos Ejemplos - Mejor Predictor Lineal (BLP)

Reemplazando lo anterior, obtenemos que:

$$\phi^2 = (\mu_{22} + \beta^2\mu_{40} - 2\beta\mu_{31})/\mu_{20}^2$$

Volviendo al análogo muestral del BLP, podemos escribir dicho término como:

$$B - \beta = (B^* - \beta) + (B - B^*)$$

Reemplazando y reordenando obtenemos que:

$$B - B^* = (1/M_{20}^*)\left[(M_{11} - M_{11}^*) - (M_{11}/M_{20})(M_{20} - M_{20}^*)\right]$$

entonces,

$$\sqrt{n}(B - B^*) = (1/M_{20}^*)\left[\sqrt{n}(M_{11} - M_{11}^*) - (M_{11}/M_{20})\sqrt{n}(M_{20} - M_{20}^*)\right]$$

Además,

$$M_{20}^* \xrightarrow{p} \mu_{20}, \quad (M_{11}/M_{20}) \xrightarrow{p} \mu_{11}/\mu_{20} \quad \text{y} \quad \sqrt{n}(M_{20} - M_{20}^*) \xrightarrow{p} 0$$

Entonces $\sqrt{n}(B - B^*) \xrightarrow{p} 0$.

Algunos Ejemplos - Mejor Predictor Lineal (BLP)

Como $\sqrt{n}(B - B^*) \xrightarrow{p} 0$ y sabemos que:

$$\sqrt{n}(B - \beta) = \sqrt{n}(B^* - \beta) + \sqrt{n}(B - B^*)$$

Por Slutsky (S3) entonces la distribución asintótica de B es la misma que de B^*. Por lo tanto:

$$B \xrightarrow{d} N(\beta, \phi^2/n)$$