Teoría Econométrica I - EAE- 250-A

Elementos de teoría asintótica

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Septiembre 2021

Introducción

- La clase pasada vimos que si suponiamos normalidad podiamos hacer inferencia exacta sobre los parámetros del modelo basado en la distribución muestral del estimador
- Ahora, levantaremos el supuesto de normalidad, lo que nos obliga a encontrar la **distribución muestral del estimador MCO**.
- Usaremos las leyes de grandes números y teoremas centrales del límite

  $\Rightarrow$ También requieren hacer supuestos pero más débiles
- Nos vamos a estar preguntando **¿Cúales son las propiedades de las secuencias de variables aleatorias y sus distribuciones en el limite?**

Convergencia en distribución

- Una secuencia de vectores aleatorios $\mathbf{x_n}$ **converge en distribución** a la distribución de $\mathbf{x}$, $\mathbf{x_n} \xrightarrow{d} \mathbf{x}$, si se cumple que:

$$\lim_{n \to \infty} F_n(x) = F(x)$$

$\forall\, x$ donde $F(x)$ es continua:

$$F_n(x) = Pr(X_n \leq x) \wedge F(x) = Pr(X \leq x)$$

- A esta función $F(x)$ se le denomina ***Limiting Distribution***, la cual no depende de n. **Cuando $n \longrightarrow \infty$ se llega a una distribución final.**
- Ahora, si $\mathbf{x_n} \xrightarrow{d} \mathbf{x}$, podemos aproximar probabilidades de $\mathbf{x_n}$ si n es grande:

$$Pr(x_n \in B) \approx \int_B dF(x)$$

Distribución en el limite

- La distribución en el límite debe satisfacer todas las condiciones de la Cumulative Distribution function (cdf).
- Si $\lim_{n\to\infty} F_n(x)$ es discontinua, puede que no sea una cdf.
- Recordamos: una cdf debe cumplir cuatro supuestos básicos:
  (i) $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$
  (ii) Monótona no decreciente.
  (iii) Continua por la derecha.
- Continua por la derecha es un concepto más débil que continuidad
- Recordamos Una función f cualquiera es continua por la derecha en un punto c si:

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 \text{ tal que } \forall\, x \in c < x < c + \delta$$

satisface:

$$|f(x) - f(c)| < \varepsilon$$

- Tip: Alcanza con que los intervalos sean abiertos por derecha. No puede haber 'saltos' si nos acercamos de

Ejemplos

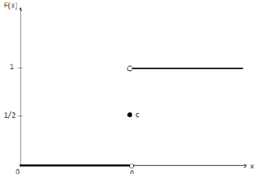
- **EJEMPLO 1:** No continua por la derecha.

$$X_n \sim N\left(0, \frac{1}{n}\right)$$

$$F_n(x) = Pr(X_n \leq x) = Pr(\sqrt{n}X_n \leq \sqrt{n}x) = \Phi(\sqrt{n}x)$$

$$F_n(x) = \Phi(\sqrt{n}x)$$

$$\lim_{n\to\infty} F_n(x) = \left\{ \begin{array}{ccc} 1 & \text{si} & x > 0 \\ \frac{1}{2} & \text{si} & x = 0 \\ 0 & si & x < 0 \end{array} \right\}$$



Convergencia en probabilidad

- Una sucesión de variable aleatoria X_n, **converge en probabilidad** a x (no estocástico) si:

$$\lim_{n \to \infty} Pr\{\|X_n - x\| > \varepsilon\} = 0, \text{ con } \varepsilon > 0$$

y se denota:

$$\operatorname{plim}_{n \to \infty} X_n = x \ \vee \ X_n \xrightarrow{p} x$$

- Corolario:

$$X_n \xrightarrow{p} x \iff \operatorname{plim}_{n \to \infty}(X_n - x) = 0$$

- La intuición de esta definición es que toda la masa de probabilidad está en una vecindad de x muy pequeña.

Convergencia en media cuadrática

- X_n **converge en media cuadrática** a x (no estocástico) si:

$$\lim_{n\to\infty} \mathbb{E}\left(\|X_n - x\|^2\right) = 0$$

- En muchos casos equivale a decir que la varianza se va contrayendo (cuando x es el valor esperado)
- Se denota:

$$X_n \xrightarrow{q.m.} x$$

  se puede demostrar que:

$$X_n \xrightarrow{q.m.} x \Longrightarrow X_n \xrightarrow{p} x$$

- Pero antes, necesitamos revisar dos desigualdades muy importantes.

Desigualdad de Markov

- Si Z (variable aleatoria) no negativa, escalar:

$$Pr(Z > k) \leq \frac{E(Z)}{k} \text{, con } k > 0$$

- **Demostración** (con variable continua):

$$\begin{aligned} E(Z) = \int_0^\infty ZdF_Z(Z) &= \int_0^k ZdF_Z(Z) + \int_k^\infty ZdF_Z(Z) \\ &= \int_0^k ZdF_Z(Z) + \int_k^\infty (Z - k + k)dF_Z(Z) \\ &= \int_0^k ZdF_Z(Z) + \int_k^\infty (Z - k)dF_Z(Z) + \int_k^\infty kdF_Z(Z) \end{aligned}$$

con $\int_0^k ZdF_Z(Z) + \int_k^\infty (Z - k)dF_Z(Z) \geq 0$. Luego,

$$\begin{aligned} \mathbb{E}(Z) &\geq k \int_k^\infty dF_Z(Z) \\ \mathbb{E}(Z) &\geq k \cdot Pr(Z > k) \\ \implies Pr(Z > k) &\leq \frac{\mathbb{E}(Z)}{k} \end{aligned}$$

Desigualdad de Chebychev (otra vez!!)

- Usando la desigualdad de markov podemos llegar a la de Chebychev
- Remplazamos $Z = (X - \mathbb{E}(X))^2$ para una variable aleatoria escalar X (con segundo momento finito)
- y remplazamos $k = \varepsilon^2$, $\forall\, \varepsilon > 0$. Aplicando la *desigualdad de Markov*:

$$\begin{aligned} Pr\left\{|X_n - \mathbb{E}(X)| > \varepsilon\right\} &= Pr\left\{[X - \mathbb{E}(x)]^2 > \varepsilon^2\right\} \\ Pr\left\{|X_n - \mathbb{E}(X)| > \varepsilon\right\} &\leq \frac{\mathbb{E}\left[(X - \mathbb{E}(X))^2\right]}{\varepsilon^2} \end{aligned}$$

Usando ambas desigualdades podemos demostrar que **convergencia en media cuadrática implica convergencia en probabilidad**, es decir,

$$X_n \xrightarrow{q.m.} x \Longrightarrow X_n \xrightarrow{p} x$$

Demostración

1. Convergencia en media cuadrática $\Longrightarrow \lim_{n\to\infty} \mathbb{E}\left\{\|X_n - x\|^2\right\} = 0$
2. Por *desigualdad de Markov*:

$$Pr\left\{\|X_n - x\| > \varepsilon\right\} \leq \frac{\mathbb{E}\left\{\|X_n - x\|^2\right\}}{\varepsilon^2}, \text{ con } \varepsilon > 0$$

3. Aplicamos limites

$$\lim_{n\to\infty} Pr\left\{\|X_n - x\| > \varepsilon\right\} \leq 0$$

4. Usamos definición de probabilidad:

$$\lim_{n\to\infty} Pr\left\{\|X_n - x\| > \varepsilon\right\} = 0$$

Por lo tanto, convergencia en media cuadrática es más fuerte que convergencia en probabilidad.

Convergencia Almost Surely

- X_n converge almost surely a x (no estocástico) si

$$Pr\left\{\lim_{n\to\infty} X_n = x\right\} = 1$$

- Se denota:

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

- La **convergencia Almost Surely implica convergencia en probabilidad**.

Ley Débil de los grandes Números (WLLN) de Khintchine

- Sea X_n una secuencia de variables aleatorias iid
- Sea x_i escalar iid y se cumple:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

con

$$\mathbb{E}(x_i) = \mu \text{ (finito)}$$

entonces:

$$\overline{X}_n \xrightarrow{p} E(x_i) = \mu$$

- Esto nos dice que el proceso generador de datos arroja realizaciones de igual distribución

Ley Débil de los grandes Números (WLLN) de Chebychev

- Sea X_n una secuencia de variables aleatorias independientes no idénticas
- Sea x_i escalar y se cumple: Si X_n es escalar (variable aleatoria) con:

$$E(X_n) = \mu_n \ , \ V(X_n) = \sigma_n^2 \ , \ Cov(X_i X_j) = 0$$

  (se relajó el supuesto IID) y además:

$$\frac{1}{n}\Sigma_{i=1}^{n}\sigma_i^2 < M$$

  entonces:

$$\bar{X}_n - E(\bar{X}_n) \xrightarrow{p} 0$$

- Las leyes débiles de los grandes nos dicen que para cualquier distribución marginal distinta de cero, no importa que tan cercana, hay una alta probabilidad de que la media muestral de un numero suficiente grandes de observaciones esta cera al media poblacional de la distribución marginal
- Las leyes débiles hablan sobre la convergencia de una secuencia de probabilidades; la secuencia de variables aleatorias convergen en probabilidad a la media poblacional μ cuando n es lo suficientemente grande.

Ley de los grandes números (Law of Large Numbers)

- Sea $X_i \sim iid$ con $E(X_i) = \mu$ y $var(X_i) = \sigma^2$
- Entonces

$$Pr\left\{\lim_{n\to\infty} (\bar{X}_n - \mu) \leq \varepsilon\right\} = 1$$

- La ley ‘fuerte’ de los grandes números nos dice que con probabilidad 1 la secuencia de medias muéstrales converge a medida que el tamaño muestral crece a una constante μ que el la esperanza poblacional de la variable aleatoria
- Asociado a la idea de converge almost surely

Teorema Central del Límite (CLT)

- ***Teorema Central del Límite (CLT) de Lindeberg-Levy***.
- Sea x_i una variable aleatoria escalar IID, con:

$$\mathbb{E}(x_i) = \mu \text{ , } V(x_i) = \sigma^2 \text{ finitos, tenemos que:}$$

$$Z_n = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

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

- ***Teorema Central de Límite (CLT) de Lindeberg-Levy Multivariado***.
  Sea $\bar{\mathbf{X}}_\mathbf{n}$ el promedio muestral de vectores $\{\mathbf{x_i}\}$ con $E(\mathbf{x_i}) = \mu$ (vector) y $V(\mathbf{x_i}) = \Sigma$.

$$\sqrt{n}(\bar{\mathbf{X}}_\mathbf{n} - \mu) \xrightarrow{d} N(0,\Sigma)$$

  Notar que esta distribución en el límite no puede tener n adentro

Teorema de continuidad

- Algunos autores como Amemiya llaman a este teorema como **continuous mapping theorem**
- El nombre no importa mucho. Lo importante es lo que sigue.
- Sea X_n vector aleatorio tal que $X_n \xrightarrow{p} x_0$ y una función $g(x)$ continua para $x = x_0$, entonces

$$g(X_n) \xrightarrow{p} g(x_0).$$

Demostración

- Por definición de continuidad sabemos que: $\forall \epsilon > 0 \quad \exists \delta > 0$ tal que si
$$\|X_n - x_0\| < \delta \Rightarrow \|g(X_n) - g(x_0)\| < \epsilon.$$
- Ademas sabemos que si tenemos dos eventos A y B tales que $A \Rightarrow B$, entonces $Pr(A) \leq Pr(B)$.
- Luego si tomamos
$$A = \{ \|X_n - x_0\| < \delta \}$$
$$B = \{ \|g(X_n) - g(x_0)\| < \epsilon \}$$
- Luego
$$\begin{aligned} & Pr\{ \|X_n - x_0\| < \delta \} \leq Pr\{ \|g(X_n) - g(x_0)\| < \epsilon \} \\ \Rightarrow \quad & \lim_{n\to\infty} Pr\{ \| \|X_n - x_0\| < \delta \} \leq \lim_{n\to\infty} Pr\{ \|g(X_n) - g(x_0)\| < \epsilon \} \\ & 1 \leq \lim_{n\to\infty} Pr\{ \|g(X_n) - g(x_0)\| < \epsilon \} \\ \Rightarrow \quad & \lim_{n\to\infty} Pr\{ \|g(X_n) - g(x_0)\| < \epsilon \} = 1 \end{aligned}$$
- Concluimos entonces que $g(X_n) \xrightarrow{p} g(x_0)$.$\blacksquare$

Teorema de Slutzky

- Si los vectores aleatorios X_n e Y_n tienen la misma dimensión,
- Y además se tiene que $X_n \xrightarrow{p} x_0$, $Y_n \xrightarrow{d} Y$ y c una constante:
- Entonces:

(i) $c \xrightarrow{d} c$

(ii) $X_n + Y_n \xrightarrow{d} x_0 + Y$

(iii) $X_n' Y_n \xrightarrow{d} x_0' Y$

(iv) $X_n * Y_n^{-1} \xrightarrow{d} x_0 / y_0$ (siempre que Y_n^{-1} exista y sea comformable con X_n y $y_0 \neq 0$)

Teorema del mapeo continuo/Mann-Wald

1. Si $X_n \xrightarrow{d} X$ y $g(x)$ es continua para todo x, entonces

$$g(X_n) \xrightarrow{d} g(X).$$

2. Si $W_n = X_n + Z_n$ con $X_N \xrightarrow{d} X$ y $Z_n \xrightarrow{p} 0$, entonces: $W_n \xrightarrow{d} X$

   $\Rightarrow$ ***Corolary:*** si $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

- **EJEMPLO:** Sea $Z_n \equiv \frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sigma} \xrightarrow{d} N(0, I)$,

  Queremos saber la distribución de $T = Z_n' Z_n = n(\hat{\theta} - \theta_0)'(\hat{\theta} - \theta_0)$.

  Por Teorema de Mapeo Continuo tenemos que $T \xrightarrow{d} \chi^2_{dim(\hat{\theta})}$.

Otros resultados útiles en convergencias en distribución

1. Si $A_n \xrightarrow{p} 0$ y $W_n \xrightarrow{p} W \Rightarrow A_n W_n \xrightarrow{p} 0$
2. (Cramer) Si $A_n \xrightarrow{p} A$ y $X_n \xrightarrow{d} X \Rightarrow A_n X_n \xrightarrow{d} AX$

Método delta

- Sea $\hat{\theta}_n$ un vector aleatorio asintóticamente normal con:

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

- Y sea $g(\theta)$ una función continuamente diferenciable en $\theta = \theta_0$, con Jacobiano

$$G_0 \equiv \left.\frac{\partial\, g(\theta)}{\partial\, \theta}\right|_{\theta=\theta_0},$$

- Entonces

$$\sqrt{n}\,(g(\hat{\theta}) - g(\theta_0)) \xrightarrow{d} N(0, G_0\, \Sigma\, G_0').$$

- ***Versión Factible:*** Si $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$ y $\hat{\sigma^2}_n \xrightarrow{p} \sigma$:

$$\frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\sigma}_n} \xrightarrow{d} N(0, 1)$$

Demostración del Delta Method

- De acuerdo con el **Teorema del Valor Medio** sabemos que $\exists\, \theta_n^* \in [\theta_0\, ,\, \hat{\theta}_n]$ tal que

$$\left(g(\hat{\theta}_n) - g(\theta_0)\right) = \left.\frac{\partial\, g(\theta)}{\partial\, \theta}\right|_{\theta=\theta_n^*} \left(\hat{\theta}_n - \theta_0\right)$$

$$\sqrt{n}\,\left(g(\hat{\theta}_n) - g(\theta_0)\right) = \left.\frac{\partial\, g(\theta)}{\partial\, \theta}\right|_{\theta=\theta_n^*} \sqrt{n}\,\left(\hat{\theta}_n - \theta_0\right).$$

Dado que $\hat{\theta}_n \xrightarrow{p} \theta_0$, entonces $\theta_n^* \xrightarrow{p} \theta_0$. Luego por Teorema de Continuidad

$$\left.\frac{\partial\, g(\theta)}{\partial\, \theta}\right|_{\theta=\theta_n^*} \xrightarrow{p} G_0.$$

Además por hipótesis tenemos que $\sqrt{n}\,\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, \Sigma)$.
Luego por Teorema de Slutsky

$$\sqrt{n}\,\left(g(\hat{\theta}_n) - g(\theta_0)\right) = \left.\frac{\partial\, g(\theta)}{\partial\, \theta}\right|_{\theta=\theta_n^*} \sqrt{n}\,\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} G_0\, N(0, \Sigma) = N(0, G_0 \Sigma\, G$$

Propiedades asintóticas del MCO

- Vamos a relajar algunos supuestos:
  (i) Normalidad; (ii) Homocedasticidad; (ii) No Estocasticidad de $X \Rightarrow$ Ya relajada
- Recordemos que

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

Tomando $y_i = {x_i}'\beta + \epsilon_i$, donde x_i es el i-ésimo vector fila de X.

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i'x_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i'(x_i\beta + \epsilon_i)\right)$$

$$\hat{\beta} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i'x_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i\right)$$

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i'x_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\epsilon_i\right) \quad (1)$$

Consistencia del estimador MCO

- **Definición:** Un estimador $\hat{\theta}$ es consistente si converge en probabilidad al verdadero valor del parametro

$$\mathrm{plim}_{n\to\infty}\hat{\theta} = \theta$$

- Para $\hat{\beta}_{MCO}$ tenemos

$$\hat{\beta} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i\right)}_{\text{by LLN} \xrightarrow{p} \mathbb{E}(x_i\epsilon_i)=0} \xrightarrow{p} \beta$$

Ya que

$$\begin{aligned}
\mathbb{E}(x_i\epsilon_i|x_i) &= \mathbb{E}(x_i\,(y_i - x_i\beta)) \\
&= \mathbb{E}(x_i' y_i) - \mathbb{E}(x_i' x_i)\beta \\
&= \mathbb{E}(x_i' y_i) - \mathbb{E}(x_i' x_i)(\mathbb{E}(x_i' x_i))^{-1}\mathbb{E}(x_i' y_i) \\
&= \mathbb{E}(x_i' y_i) - \mathbb{E}(x_i' y_i) \\
&= 0
\end{aligned}$$

Propiedades asintóticas del MCO

- Por WLLN
$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i \xrightarrow{p} E(x_i' x_i) = D$$
Luego por Teorema de Continuidad tenemos que
$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \xrightarrow{p} D^{-1}.$$
Notemos que
$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i = \frac{\sqrt{n}}{n}\sum_{i=1}^{n} x_i' \epsilon_i$$
por lo que podemos intentar aplicar el Teorema Central de Límite de Lindeberg-Lèvy (CLT L-L),

Propiedades asintóticas del MCO

- Para aplicar el TCL recordemos que:

$$
\begin{aligned}
\mathbb{E}(x_i \epsilon_i | x_i) &= \mathbb{E}(x_i\,(y_i - x_i\beta)) \\
&= \mathbb{E}(x_i' y_i) - \mathbb{E}(x_i' x_i)\beta \\
&= \mathbb{E}(x_i' y_i) - \mathbb{E}(x_i' x_i)(\mathbb{E}(x_i' x_i))^{-1}\mathbb{E}(x_i' y_i) \\
&= \mathbb{E}(x_i' y_i) - \mathbb{E}(x_i' y_i) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
V(x_i' \epsilon_i | x_i) &= \mathbb{E}(x_i' \epsilon_i \epsilon_i' x_i) \\
&= \mathbb{E}(\epsilon_i^2 x_i' x_i) \\
&= C
\end{aligned}
$$

Propiedades asintóticas del MCO

- Luego asumiendo proceso *i.i.d*, primer y segundo momento finitos, por TCL L-L

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} N(0, C).$$

- Aplicando el Teorema de Slutsky tenemos que

$$\sqrt{n}\,(\hat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i' \epsilon_i\right) \xrightarrow{d} D^{-1}N(0, C)$$

Donde:

$$D^{-1}N(0, C) = N(0, \underbrace{D^{-1}C\,D^{-1}}_{\text{Varianza Asintótica}})$$

Estimador Consistente de Matriz de Varianzas y Covarianzas

- El estimador natural de D sería: $\hat{D} = N^{-1}\sum x_i' x_i \xrightarrow{p} D$ debido a que x es una variable observable y conocida. En donde x_i es el i-ésimo vector fila.
- Para tener un estimador C es un poco más complicado por que no observamos ϵ y precisamos usar una estimación, es decir $\hat{\epsilon}$
- Usando un criterio de analogía entre población y muestra, podemos plantear:

$$\begin{aligned}
\hat{C} &= N^{-1}\sum \hat{\epsilon}_i^2 x_i' x_i \\
\hat{C} &= N^{-1}\sum (y_i - x_i\hat{\beta})^2 x_i' x_i \\
\hat{C} &= N^{-1}\sum (\underbrace{(y_i - x_i\beta)}_{\epsilon} + x_i(\beta - \hat{\beta}))^2 x_i' x_i \\
\hat{C} &= N^{-1}\sum \epsilon_i^2 x_i' x_i + N^{-1}\sum (x_i(\beta - \hat{\beta}))^2 x_i' x_i \\
&= +2N^{-1}\sum (y_i - x_i\beta)(x_i(\beta - \hat{\beta}))x_i' x_i
\end{aligned}$$

Estimador Consistente de Matriz de Varianzas y Covarianzas

- Donde

$$\begin{aligned} \hat{C} &= \underbrace{N^{-1}\sum \epsilon_i^2 x_i' x_i}_{porWLLN \xrightarrow{p} \mathbb{E}(\epsilon_i^2 x_i' x_i) = C} + \underbrace{N^{-1}\sum (x_i(\beta - \hat{\beta}))^2 x_i' x_i}_{\xrightarrow{p} 0 \text{ por } (\hat{\beta} \xrightarrow{p} \beta)} \\ &= + \underbrace{2N^{-1}\sum (y_i - x_i\beta)(x_i(\beta - \hat{\beta}))x_i' x_i}_{\xrightarrow{p} 0 \text{ por } (\hat{\beta} \xrightarrow{p} \beta)} \end{aligned}$$

- Tenemos entonces $\hat{D} \xrightarrow{p} D$
- Tenemos asi el **estimador de Eicker-Huwber-White (E-H-W)** es:

$$\begin{aligned} [\hat{D^{-1}}\hat{C}\hat{D^{-1}}] &= V_{E-H-W}(\hat{\beta}) \\ &= [N^{-1}\sum x_i' x_i]^{-1}[N^{-1}\sum \hat{\epsilon_i^2} x_i' x_i][N^{-1}\sum x_i' x_i]^{-1} \end{aligned}$$

- Es **robusto a heteroscedasticidad** ya que en ningún momento asumimos que $\sigma_i = \sigma$

Overview

Overview:

2 convergence concepts:	In probability	In distribution
Basic" results for sample means:	LLN $\overline{X}_n \xrightarrow{p} \mu$	CLT $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$
Extensions (functions of sample means)	Slutsky theorem $g(\overline{X}_n) \xrightarrow{p} g(\mu)$	Delta method $\sqrt{n}\left[g(\overline{X}_n) - g(\mu)\right] \xrightarrow{d}$ $\xrightarrow{d} N(0, \sigma^2 g'(\mu)^2)$

Figure: *Overview de los conceptos de la teoría asíntotica*

Inferencia

- Todos estos conceptos nos van a permitir hacer inferencia, incluso no lineal!

  Test de Wald Generalizado

- El test de Wald que teníamos se planteaba como :

$$H_0 : R'\beta = r$$
$$H_1 : R'\beta \neq r$$

- En este modelo no podíamos realizar inferencia sobre funciones de los estimadores que siguieran esta forma:

$$H_0 : g(\beta) = g(\beta_0)$$
$$H_1 : g(\beta) \neq g(\beta_0)$$

  Donde $g(\beta) : \mathcal{R}^k \rightarrow \mathcal{R}$ es una función continua y diferenciable

- Definimos $G = \frac{\partial g(\beta)}{\partial \beta'}$ como el Jacobiano de la función $g(\beta)$

  $\Rightarrow$ El Jacobiano es de rango completo, lo que equivale a la independencia de las restricciones en el modelo lineal.

Inferencia

- Usando el método delta y el resultado de la ecuación (1) tenemos, bajo la nula, el siguiente resultado:

$$\sqrt{N}(g(\hat{\beta}) - g(\beta_0)) \xrightarrow{\text{d}} N(0, G[D^{-1}CD^{-1}]G')$$

- Este resultados nos permite construir un test de Wald ya que tenemos a distribución de $g(\hat{\beta})$
- El teorema de mapeo continuo nos permite derivar la distribución del test de Wald a una χ^2
- Dado que $\hat{G} = \frac{\partial g(\beta)}{\partial \beta'}|_{\hat{\beta}} \xrightarrow{\text{p}} G$ por WLLN y el teorema de la continuidad. Con esto el **Test de Wald generalizado** es:

$$\boxed{GW_N = N(g(\hat{\beta}) - g(\beta_0))^T[\hat{G}[\hat{D}^{-1}\hat{C}\hat{D}^{-1}]\hat{G}^T]^{-1}(g(\hat{\beta}) - g(\beta_0)) \xrightarrow{\text{d}} \chi^2_{Rango[g(\hat{\beta})]}}$$