ESTATÍSTICA AVANÇADA: MODELOS NÃO LINEARES

Modulo 3: Teoria Assintótica

Gilvan Guedes, Cedeplar - UFMG
Melissa Pinho, Estatística - UFMG

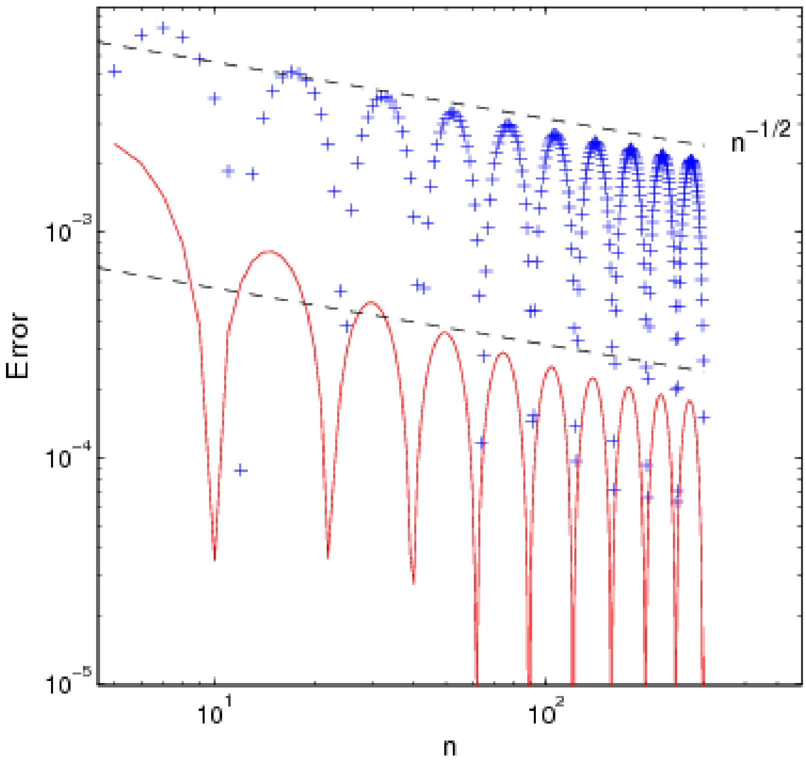
Escola do Legislativo - ALMG
Belo Horizonte, Minas Gerais

21 de setembro de 2015

Sumário

1 Propriedades Assintóticas dos Estimadores

```
Esse texto é de autoria da Profa. Sueli Moro (Departamento de
Economia/UFMG) e está estritamente proibida sua reprodução. A maior
parte das fórmulas é baseada no livro do (Greene, 2012). Algumas
pequenas alterações no texto original foram feitas para adequar linguagem
e notação para o LateX.
```



A econometria tem se tornado "cheia de assintótico"(Leamer, 1988).

A teoria assintótica trata d**o que acontece a uma estatística ou estimador quando o tamanho da amostra se torna muito grande**. Essa estatística pode ser a média, variância, os coeficientes do modelo linear e até mesmo os testes estatísticos.

Indexamos sempre em n ou T para mostrar o papel do tamanho da amostra:

$$\hat{\beta}_n \text{ ou } \hat{\beta}_t$$

O quão grande precisa ser uma amostra para que o estimador mostre as suas propriedades assintóticas?

Goldfeld and Quandat (1972) mostram um exemplo em que uma amostra de tamanho 30 já é suficientemente grande. Mostram também outro exemplo em que é necessária uma amostra de 200 observações. Segundo os autores, **grandes amostras são mais importantes quando existe maior interesse na variância do que nos valores**.

Como quase sempre trabalhamos com pequenas amostras (finitas), defender um estimador com base nas suas propriedades assintóticas faria sentido somente se **estimadores com melhores propriedades assintóticas também tivessem melhores propriedades em amostras finitas**. E parece que é isso que acontece!

2 Propriedades Assintóticas

Um **estimador não viesado** é aquele cuja distribuição amostral é centrada no verdadeiro valor do parâmetro; ou seja:

$$E[\hat{\theta}] = \theta$$

Essa propriedade não depende do tamanho da amostra: um estimador não viesado é não viesado em pequenas e grandes amostras!

Mas às vezes não é possível encontrar um estimador que tenha boas propriedades em pequenas amostras. Quando isso acontece, precisamos justificar o uso do estimador com base em suas propriedades assintóticas, ou seja, a sua distribuição em grandes amostras.

A distribuição de um estimador muda com o tamanho da amostra. A média da amostra, por exemplo, tem uma distribuição amostral que é centrada na média da população:

$$E[\bar{X}] = \mu$$

Mas **a variância da média torna-se menor à medida que a amostra aumenta**:

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Quais são as propriedades assintóticas que precisamos definir para um estimador?

1. **Consistência**: utiliza o conceito de **convergência em probabilidade**. Questão: a distribuição de $\hat{\beta}$ colapsa para um valor específico quando n aumenta?

2. **Eficiência assintótica**: utiliza o conceito de **convergência em distribuição** e está relacionada com o conceito de distribuição assintótica. Questão: a distribuição de $\hat{\beta}$ se aproxima de uma forma conhecida (ex. normal) quando n aumenta?

Para provar que um estimador é eficiente assintoticamente, é necessário conhecer a distribuição assintótica desse estimador! **Observação:** ela precisa ser normal!

Mas **o que é a distribuição assintótica de um estimador?**

Considerar a distribuição de um estimador $\hat{\beta}$ para amostras cada vez maiores: **se as distribuições vão se tornando cada vez mais parecidas com alguma distribuição específica, esta é chamada de distribuição assintótica** desse estimador!

Algumas considerações:

1. Se a distribuição assintótica de $\hat{\beta}$ se concentra em um valor k quando a amostra tende ao infinito, diz-se que k é a **probabilidade limite** de $\hat{\beta}$. Se $\lim_{n\to\infty}\hat{\beta} = \beta$, dizemos que $\hat{\beta}$ é um estimador consistente!

2. A variância da distribuição assintótica é chamada de variância assintótica.

Então, se $\hat{\beta}$ é consistente e a sua variância assintótica é menor que a variância assintótica de qualquer outro estimador consistente, $\hat{\beta}$ é chamado de **assintoticamente eficiente**!

- A propriedade $plim\hat{\beta} = \beta$ pode ser vista como o equivalente de ausência de viés em grandes amostras (**atenção**, isso somente no sentido de que se refere ao valor do parâmetro)

- Consistência pode ser pensada como o equivalente da propriedade de *minimum mean square error* (mínimo erro quadrático médio)

que MSE pode ser definido como:

$$MSE = \text{Var}(\hat{\theta}) + [\text{Viés}(\hat{\theta})]^2$$

A variância de um estimador assintoticamente eficiente tende a zero mais rapidamente que a variância de qualquer outro estimador consistente.

2.1 Consistência de um estimador

Observação: mesma intuição da esperança, no sentido que se refere a valores.

Para estabelecer consistência, precisamos do conceito de **convergência em probabilidade**. O conceito de convergência em probabilidade é **usado para definir a consistência de estimadores**.

Convergência em Probabilidade:

A sequência de variáveis aleatórias

$$z_1, z_2, \ldots, z_n$$

converge em probabilidade para a variável aleatória z se:

$$\lim_{T\to\infty} Prob[|z_n - z| > \epsilon] = 0 \ \forall \ \epsilon > 0$$

Diz-se, então que:

$$z_n \xrightarrow{p} z$$

Ou seja, z_n converge em probabilidade para z. E que:

$$plim(z_n) = z$$

Isso quer dizer que a **probabilidade da diferença ser maior que qualquer número positivo**, por menor que ele seja, **é zero**.

Uma **condição suficiente (mas não necessária)** para que ocorra a convergência é a **convergência em quadrado médio ou média quadrática** (*mean square*). É condição suficiente porque pode haver convergência sem que ocorra convergência em média quadrática.

Exemplo Uma variável aleatória z toma dois valores 0 e T com probabilidades:

$$P(z = 0) = 1 - \frac{1}{T}$$

e

$$P(z = T) = \frac{1}{T}$$

Quando T aumenta a $P(z = 0)$ tende a 1 e a $P(z = T)$ tende a 0. Nesse caso, z converge em probabilidade a 0 e, portanto:

$$\lim_{T \to +\infty} z_T = 0$$

e

$$z_T \xrightarrow{p} 0 \text{ ou seja, } plim(z_T) = 0$$

2.2 Convergência em quadrado médio ou média quadrática

Se z_n tem média μ_n e variância σ_n^2 tais que:

$$\lim_{n \to +\infty} E[z_n] = c$$

e

$$\lim_{n \to +\infty} Var[z_n] = 0$$

então z_n converge em média quadrática para c e

$$plim(z_n) = c$$

A prova é obtida pelo Teorema de Chebychev. Ou seja, limite da esperança é uma constante e limite da variância é zero.

Para nosso estimador $\hat{\beta}$, teriamos:

$$\lim_{n \to +\infty} E[\hat{\beta}_n] = \beta$$

e

$$\lim_{n \to +\infty} Var[\hat{\beta}_n] = 0$$

Convergência em média quadrática implica em convergência em probabilidade, mas a recíproca não é verdadeira!

Observação: Convergência em média quadrática: o limite da esperança tem que ser uma constante e o limite da variância tem que ser zero.

Exemplo de um caso em que a convergência em probabilidade não implica em convergência em quadrado médio: Seja $x_n = 0, T$, $P(x_n = 0) = 1 - \frac{1}{T}$ e $P(x_n = T) = \frac{1}{T}$. Nesse caso:

$$E[x_n] = 0\left(1 - \frac{1}{T}\right) + T\left(\frac{1}{T}\right) = 1$$

e

$$\lim_{n\to+\infty} E[x_n] = 1$$

mas a probabilidade limite é zero, porque:

$$x_n \xrightarrow{p} 0$$

ou seja,

$$plim(x_n) = 0$$

Conclusão: O limite da esperança é 1 e a *plim* é 0.

O limite da variância também não é zero:

$$E(x_n^2) = (0)^2\left(1 - \tfrac{1}{T}\right) + (T)^2\tfrac{1}{T} = T$$

$$\begin{gathered} Var(x_n) = E[x_n - E(x_n)]^2 \\ Var(x_n) = E[x_n - 1]^2 = E[x_n^2] - 2E[x_n] + (1)^2 \\ Var(x_n) = T - 2 + 1 = T - 1 \end{gathered} \tag{1}$$

Conclusão: A variável z_n converge em probabilidade, mas não converge em quadrado médio, porque **a média converge para uma constante, mas a variância não converge para zero**.

Felizmente, isso é raro de acontecer. Normalmente, convergência em probabilidade implica em convergência em quadrado médio.

O conceito de convergência em probabilidade é usado para definir a consistência de estimadores:

- **Convergência** $\rightarrow$ Para onde vai? *plim*
- **Consistência** $\rightarrow$ Vai para onde eu quero que vá e a variância se aproxima de zero? $plim\hat{\beta} = \beta$?

Então, um estimador será consistente se $plim\hat{\beta} = \beta$ **e sua variância se aproxima de zero quando** n **aumenta**.

Consistência da média de uma amostra

A média de uma amostra é um estimador consistente para a média da população. Isto porque:

$$E[\bar{x}] = \mu$$

e

$$Var[\bar{x}] = \frac{\sigma^2}{n}$$

Assim,

$$\lim_{n \to +\infty} var(\bar{x}) = 0$$

e

$$\lim_{n \to +\infty} E[\bar{x}] = \mu$$

2.3 O Teorema de Slutsky

É uma consequência interessante da consistência. É uma das razões que possibilita usar a teoria assintótica. Isso porque a álgebra associada ao cálculo das esperanças, que é necessária para encontrar as propriedades em amostras finitas, ás vezes é muito difícil, pode ser até impossível, e no limite tudo fica mais fácil!

Teorema de Slutsky:

Para uma função contínua de x_n, $g(x_n)$,

$$plim g(x_n) = g[plim(x_n)]$$

Observação: $g(x_n)$ precisa ser uma função contínua. Isto é: a *plim* de uma função contínua é a função avaliada na *plim*.

Isso não é verdade para a esperança matemática, pois o valor esperado de uma função não-linear de uma estatística não é uma função não-linear do valor esperado daquela estatística.

Exemplo Seja g uma função quadrática de um estimador $g(b) = b^2$, então:

$$E[b^2] = Var(b) + [E(b)]^2$$

Lembre-se que $Var(x) = E(x^2) - [E(x)]^2 \to E(x^2) = Var(x) + [E(x)]^2$

Sabemos que $E[b^2] \neq [E(b)]^2$. Mas no limite $Var(b)$ tende a zero e, portanto:

$$\lim_{n \to +\infty} E[b^2] = [E(b)]^2$$

e

$$plim[b^2] = plim[E(b)]^2$$

O Teorema de Slutsky é utilizado em várias situações como, por exemplo, em *plim* da multiplicação de funções, *plim* da soma de funções, etc.

2.4 Convergência em Distribuição

Convergência em distribuição é a forma mais fraca de convergência. O Teorema Central do Limite é um teorema sobre convergência em distribuição.

Convergência em probabilidade não é a mesma coisa que convergência em distribuição. Vejamos a diferença:

- **Convergência em probabilidade:** diz que a variável aleatória converge para um valor conhecido (o parâmetro na população).
- **Convergência em distribuição:** diz que as variáveis se comportam da mesma maneira, ou seja, tem a mesma distribuição, mas não têm necessariamente os mesmos parâmetros.

É importante para a distribuição limite e para a distribuição assintótica.

Observação: Convergência em distribuição é a forma de convergência mais fraca (*weak convergence*) e não implica nenhum outro tipo de convergência. Mas todos os outros tipos de convergência implicam em convergência em distribuição, como a convergência em probabilidade.

3 Consistência de $\hat{\beta}$

Para que $\hat{\beta}$ seja um estimador consistente de β é preciso que $plim\hat{\beta} = \beta$ (o viés e a variância se aproximem de 0)!
Observação: Lembre-se que $\hat{\beta}$ é uma variável aleatória em amostras repetidas.

Se $\hat{\beta}$ convergir em probabilidade para uma constante, ou seja, a média na população, e se a sua variância convergir para zero, teremos convergência em *mean square*, e a probabilidade limite de $\hat{\beta}$ é β. Isso porque convergência em *mean square* implica em convergência em probabilidade.

Então, para onde converge $\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}$?

A partir de agora fazemos toda a análise indexada em n, ou seja, o tamanho da amostra agora é importante, porque desejamos saber as propriedades do estimador quando n aumenta.

Quando dividimos por n temos médias e queremos saber se essas médias de amostras convergem para médias na população.

3.1 Prova de Consistência de $\hat{\beta}$

O método dos Mínimos Quadrados Ordinários busca minimizar a soma dos quadrados dos erros, utilizando como base as equações normais:

$$\begin{gathered}(\beta) = \epsilon^{\mathbf{T}}\epsilon = (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) \\ S(\beta) = \mathbf{Y^TY} - \beta^{\mathbf{T}}\mathbf{X^TY} - \mathbf{Y^TX}\beta + \beta^{\mathbf{T}}\mathbf{X^TX}\beta \\ S(\beta) = \mathbf{Y^TY} - 2\beta^{\mathbf{T}}\mathbf{X^TY} + \beta^{\mathbf{T}}\mathbf{X^TX}\beta \\ \tfrac{\partial S(\beta)}{\partial\beta} = -2\mathbf{X^TY} + 2\mathbf{X^TX}\beta\end{gathered} \tag{2}$$

As equações normais são obtidas como:

$$\begin{gathered}S(\beta)/\beta = 0 \\ \tfrac{\partial S(\beta)}{\partial\beta} = -2\mathbf{X^TY} + 2\mathbf{X^TX}\beta = 0 \\ \mathbf{X^TX}\beta = \mathbf{X^TY}\end{gathered} \tag{3}$$

Como:

$$\begin{gathered}\partial S(\beta)\partial\beta = -2\mathbf{X^TY} + 2\mathbf{X^TX}\beta = 0 \\ \beta = (\mathbf{X^TX})^{-}\mathbf{1}\mathbf{X^TY} \\ \tfrac{\partial 2S(\beta)}{\partial\beta\partial\beta} = 2\mathbf{X^TX}\end{gathered} \tag{4}$$

Assim, $\mathbf{X^TX}$ é uma matriz positiva e definida.

O estimador MQO de β é:

$$\hat{\beta} = (\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X^Ty}$$

Substituindo $\mathbf{y} = \mathbf{X}\beta + \epsilon$, temos:

$$\begin{gathered}\beta = (\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X^T}(\mathbf{X}\beta + \epsilon) \\ \hat{\beta} = (\mathbf{X^TX})^{-\mathbf{1}}(\mathbf{X^TX})\beta + (\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X}\epsilon \\ \hat{\beta} = \beta + (\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X}\epsilon\end{gathered} \tag{5}$$

Multiplicando e dividindo a parte destacada de $\hat{\beta}$ por n teremos:

$$plim(\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X}\epsilon = plim\left[\frac{(\mathbf{X^TX})^{-\mathbf{1}}}{n}\frac{\mathbf{X^T}\epsilon}{n}\right]$$

Utilizando o Teorema de Slutsky:

$$plim(g(x)) = g(plim(x))$$

temos:

$$plim\left[\frac{(\mathbf{X^TX})^{-1}}{n}\frac{\mathbf{X^T}\epsilon}{n}\right] = plim\frac{(\mathbf{X^TX})^{-1}}{n}plim\frac{\mathbf{X^T}\epsilon}{n}$$

Dividir a equação acima por n faz sentido porque agora $n \to +\infty$ e também porque $\mathbf{X^TX}$ é uma matriz de somas de produtos e os elementos de sua diagonal são somas de quadrados.

Se existe um intercepto, o elemento superior à direita da matriz $\mathbf{X^TX}$ será igual a n. Então, quando $n \to +\infty$, algumas dessas somas tendem a ficar muito grandes. Por isso, não faz sentido achar $plim(\mathbf{X^TX})$. Quando olhamos a $plim(\frac{\mathbf{X^TX}}{n})$ estamos olhando para os valores médios de $\mathbf{X^TX}$, e esses são finitos quando n aumenta.

Pressuposto:

$$\lim_{n\to+\infty}\frac{\mathbf{X^TX}}{n} = \mathbf{Q}$$

em que **Q é uma matriz finita, positiva definida e não-singular**. É uma média do somatório de produtos e dos somatórios de quadrados de variáveis.

Isso implica que quando n aumenta os elementos de $\mathbf{X^TX}$ não aumentam a uma taxa maior do que n. Mas quando essa expressão se torna infinita (no caso da equação de tendência) a sua inversa se torna 0.
Vamos mostrar que:

$$\sigma^2(\mathbf{X^TX})^{-1} = var(\beta)$$

vai se aproximando de zero quando n aumenta, ou seja:

$$\lim_{n\to\infty}\sigma^2(\mathbf{X^TX})^{-1} = var(\beta) = 0$$

A expressão $\lim_{n\to+\infty}\frac{\mathbf{X^TX}}{n} = Q$ também implica que quando n aumenta, os elementos de $\mathbf{X^TX}$ são linearmente independentes no limite, uma vez que ela é não-singular, admitindo inversa. Vejamos um exemplo de violação dessa hipótese.

Equação de tendência:

$$y = \beta_1 + \beta_2 + \epsilon$$

$$\mathbf{X^TX} = \begin{bmatrix} T & \frac{T(T+1)}{2} \\ T(T+1) & \frac{T(T+1)(2T+1)}{6} \end{bmatrix}$$

Veja que a equação viola o pressuposto, pois

$$\lim_{T\to+\infty}\left(\frac{\mathbf{X^TX}}{T}\right) = \begin{bmatrix} 1 & +\infty \\ +\infty & +\infty \end{bmatrix}$$

esta matriz não é finita, ou seja, “explode”no infinito.
Outra equação que viola o pressuposto é:

$$y_t = \beta_1 + \beta_2\lambda^t + \epsilon_t$$

em que: $|\lambda| < 1$

Nesse caso,

$$\lim_{T\to+\infty}\left(\frac{\mathbf{X^TX}}{\mathbf{T}}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

que é uma matriz singular.

Felizmente, a matriz $\mathbf{X^TX}$ satisfaz as **condições de Grenander**, que são mais fracas e suficientes para a consitência dos $\hat{\beta}$.

`Condições de Grenander (Greene, 2012):`

G1 - Para cada coluna de $\mathbf{X}$, $\mathbf{x_k}$, se $d^2_{nk} = \mathbf{x_k^T x_k}$, então $\lim_{n\to+\infty} d^2_{nk} = +\infty$. Portanto, $\mathbf{x_k}$ não degenera a uma sequência de zeros. As somas dos quadrados continuarão a crescer na medida em que o tamanho da amostra aumenta. Nenhuma variável degenerará para uma sequência de zeros.

G2 - $\lim_{n\to+\infty} x^2_{ik}/d^2_{nk} = 0$ para todo $i = 1, \ldots, n$. Essa condição implica que nenhuma observação individual dominará $\mathbf{x_k^T x_k}$, e na medida em que $n \to +\infty$, observações individuais se tornarão menos importantes.

G3 - Seja $\mathbf{R}_n$ a matriz de correlação amostral das colunas da matriz $\mathbf{X}$, excluindo o termo constante caso haja um. Então, $\lim_{n\to+\infty} \mathbf{R}_n = \mathbf{C}$, uma matriz positiva definida. Essa condição implica que a condição de posto completo sempre será atingida. Já assumimos que $\mathbf{X}$ tem posto completo em uma amostra finita, então esse pressuposto assegura que a condição nunca será violada.

Até agora provamos que

$$\lim_{n\to+\infty} \frac{\mathbf{X^TX}}{n} = Q$$

ou que pelo menos $\mathbf{X^TX}$ satisfaz as condições de Grenander. Portanto, temos agora que:

$$\begin{aligned} \text{plim}\hat{\beta} &= \beta + plim\left(\frac{\mathbf{X^TX}}{n}\right)^{-1} plim\left(\frac{\mathbf{X^T}\epsilon}{n}\right) \\ \text{plim}\hat{\beta} &= \beta + \mathbf{Q}^{-1} plim\left(\frac{\mathbf{X^T}\epsilon}{n}\right) \end{aligned} \tag{6}$$

Agora precisamos fazer a prova assintótica de $plim\left(\frac{\mathbf{X^T}\epsilon}{n}\right)$. Fazendo:

$$\frac{\mathbf{X^T}\epsilon}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i = \frac{1}{n}\sum_{i=1}^{n} w_i = \bar{w}$$

Assim, teremos:

$$plim\hat{\beta} = \beta + \mathbf{Q}^{-1} plim\bar{w}$$

Supondo que $\mathbf{X}$ seja não-estocástico (e portanto, não-correlacionado com o erro), temos:

Média de $\bar{w}$

$$E[\bar{w}] = \frac{1}{n}\sum_{i=1}^{n} E[w_i] = \frac{1}{n}\sum_{i=1}^{n} E[x_i, \epsilon_i] = \frac{1}{n}\sum_{i=1}^{n} x_i E(\epsilon_i) = \frac{1}{n}\mathbf{X^T}E[\epsilon] = 0$$

Variância de $\bar{w}$

$$\begin{aligned} Var(\bar{w}) &= E[\bar{w}\bar{w}^T] = E\left[\tfrac{\mathbf{X^T}\epsilon}{n}\tfrac{\epsilon^{\mathbf{T}}\mathbf{X}}{n}\right] \\ Var(\bar{w}) &= \tfrac{1}{n}\mathbf{X^T}\mathbf{E}(\epsilon\epsilon^{\mathbf{T}}|\mathbf{X})\tfrac{\mathbf{X}}{n} = \tfrac{1}{n}\sigma^2\tfrac{\mathbf{X^T}\mathbf{X}}{n} = \tfrac{\sigma^2}{n}\tfrac{\mathbf{X^T}\mathbf{X}}{n} \end{aligned} \tag{7}$$

Tomando o limite:

$$\lim_{n\to+\infty} var(\bar{w}) = \lim_{n\to+\infty}\left(\frac{\sigma^2}{n}\frac{\mathbf{X^T}\mathbf{X}}{n}\right) = 0\mathbf{Q} = 0$$

Então, uma vez que a média de $\bar{w}$ é zero (uma constante) e sua variância converge para zero, $\bar{w}$ converge em média quadrática para zero.

Assim:

$$plim(\bar{w}) = 0$$

e

$$plim\left(\frac{1}{n}\mathbf{X^T}\epsilon\right) = 0$$

Observação: Lembre-se que isso é a prova da consistência de $\hat{\beta}$ e que estamos usando convergência em probabilidade.

Então:

$$\begin{aligned} \hat{\beta} &= \beta + (\mathbf{X^T}\mathbf{X})^{-1}\mathbf{X^T}\epsilon \\ plim\hat{\beta} &= \beta + \mathbf{Q}^{-1}plim\tfrac{\mathbf{X^T}\epsilon}{n} \\ plim\hat{\beta} &= \beta + \mathbf{Q}^{-1}0 \\ plim\hat{\beta} &= \beta \end{aligned} \tag{8}$$

Conclusão: $\hat{\beta}$ é um estimador consistente no modelo de Regressão Clássico.

3.2 Normalidade Assintótica de $\hat{\beta}$ de MQO

Nessa sessão perguntamos qual é a distribuição assintótica de $\hat{\beta}$.

Vimos que em uma amostra finita não precisamos de normalidade dos erros nem dos $\hat{\beta}$ para provar a eficiência dos $\hat{\beta}$ de MQO (a prova é obtida utilizando Gauss-Markov).

Mas vimos também que só podemos afirmar que $\hat{\beta}$ tem uma distribuição normal quando ϵ tem uma distribuição normal. Ou seja, somente podemos inferir sobre a distribuição de $\hat{\beta}$ quando os erros são normais.

Agora, para derivar a distribuição assintótica de $\hat{\beta}$ precisamos assumir a matriz $\mathbf{X}$ independente dos erros (ou seja, exogeneidade). Usamos o Teorema Central do Limite e não precisamos de normalidade em amostra finita.

Distribuições assintóticas: são também caracterizadas por suas médias e variâncias como nas distribuições finitas.

Temos então a **média assintótica** e a **variância assintótica**:

$$\hat{\beta} = E[\hat{\beta}] = \beta$$

A **média assintótica é igual à esperança assintótica** que é dada por $plim\hat{\beta} = \beta$. No caso dos estimadores, como no limite temos uma constante β, a variância é zero! Então a distribuição assintótica **não** é a distribuição no limite, porque no limite é uma distribuição **degenerada**!

Observação: No limite não há distribuição, é um pico! Então a distribuição assintótica é a distribuição da jornada final, antes de colapsar para um ponto.

A distribuição de um determinado estimador é diferente (ou pode ser), à medida que o tamanho da amostra aumenta. Pode diferir não somente na média e variância, mas também na forma matemática (lembrar que a binomial é assintoticamente uma normal).

A essência do **Teorema Central do Limite** pode ser resumida como:

Quando o tamanho da amostra aumenta a distribuição da média amostral aproxima-se da normal. Diz-se então que a normal é a distribuição assintótica da média amostral.

Quando estamos interessados em saber se os momentos da amostra convergem para os momentos da população, a **Lei dos Grandes Números** (LGN) nos dá essa resposta. Veja que há uma diferençã entre a LGN e o TCL:

- LGN: momentos $\rightarrow$ média, variância, etc.
- TCL: distribuição

3.3 Prova da Normalidade Assintótica de $\hat{\beta}$

Para derivar a distribuição de $\hat{\beta}$, não precisamos que os erros sejam normais. No entanto, precisaremos de duas condições:

- Teorema Central do Limite (TCL)
- assumir que $\mathbf{X}$ é uma matriz de variáveis independentes dos erros

Observação: Fazemos a prova utilizando uma função de $\hat{\beta}$, ao invés de $\hat{\beta}$ diretamente, porque se $\hat{\beta}$ converge para β - que é uma constante - no limite não há distribuição, apenas um pico.

Lembre-se que:

$$\begin{aligned}\hat{\beta} &= \beta + \frac{\mathbf{X^T X}}{n}^{-1} \frac{\mathbf{X^T}\epsilon}{n} \\ \hat{\beta} - \beta &= \frac{\mathbf{X^T X}}{n}^{-1} \frac{\mathbf{X^T}\epsilon}{n}\end{aligned} \tag{9}$$

Multiplicando ambos os lados por $\sqrt{n}$, obtemos:

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \left[\frac{\mathbf{X^T X}}{n}^{-1} \frac{\mathbf{X^T}\epsilon}{\sqrt{n}\sqrt{n}}\right]\sqrt{n} \\ \sqrt{n}(\hat{\beta} - \beta) &= \frac{\mathbf{X^T X}}{n}^{-1} \frac{\mathbf{X^T}\epsilon}{\sqrt{n}}\end{aligned} \tag{10}$$

Temos que a distribuição limite de $\sqrt{n}(\hat{\beta} - \beta)$ é igual à distribuição limite do lado direito, ou seja:

$$plim\left[\left(\frac{\mathbf{X^T X}}{n}\right)^{-1}\left(\frac{\mathbf{X^T}\epsilon}{\sqrt{n}}\right)\right]$$

Já obtivemos a $plim\left(\frac{\mathbf{X^T X}}{n}\right)^{-1}$, que é $\mathbf{Q}^{-1}$. Mas não temos a $plim\left(\frac{\mathbf{X^T}\epsilon}{\sqrt{n}}\right)$, embora tenhamos a $plim\left(\frac{\mathbf{X^T}\epsilon}{n}\right) = 0$.

Então, qual seria a distribuição limite de $\left(\frac{\mathbf{X^T}\epsilon}{\sqrt{n}}\right)$? Veja que essa é uma distribuição que será mais lenta para convergir, pois é indexada em $\sqrt{n}$ ao invés de n:

$$\left(\frac{1}{\sqrt{n}}\right)\mathbf{X^T}\epsilon = \sqrt{n}\bar{\mathbf{w}}$$

$$\text{Lembre-se que } \bar{\mathbf{w}} = \frac{\mathbf{X^T}\epsilon}{n} = \frac{\mathbf{X^T}\epsilon}{\sqrt{n}\sqrt{n}}$$

$$\sqrt{n}\bar{\mathbf{w}} = \sqrt{n}\frac{\mathbf{X^T}\epsilon}{\sqrt{n}\sqrt{n}} = \frac{\mathbf{X^T}\epsilon}{\sqrt{n}} = \frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon(11)$$

Calculando a esperança e a variância de $\sqrt{n}\bar{\mathbf{w}}$, temos:

$$\begin{gathered}
\text{Média:}\\
E[\sqrt{n}\bar{\mathbf{w}}] = \sqrt{n}E[\bar{\mathbf{w}}] = 0\\
\\
\text{Variância:}\\
Var[\sqrt{n}\bar{\mathbf{w}}] = nVar(\bar{\mathbf{w}})\\
\\
Var[\sqrt{n}\bar{\mathbf{w}}] = n\left[\frac{\sigma^2}{n}\frac{\mathbf{X^T X}}{n}\right]\\
\\
Var[\sqrt{n}\bar{\mathbf{w}}] = \sigma^2\frac{\mathbf{X^T X}}{n}
\end{gathered} \tag{12}$$

Se tomarmos o limite da variância de $\sqrt{n}\bar{\mathbf{w}}$, teremos:

$$\lim_{n\to+\infty} Var[\sqrt{n}\bar{\mathbf{w}}] = \lim_{n\to+\infty} \sigma^2\frac{\mathbf{X^T X}}{n} = \sigma^2\mathbf{Q}$$

Resta agora aplicar uma variante do Teorema Central do Limite (variante de Lindberg-Feller) ao vetor $\sqrt{n}\bar{\mathbf{w}}$. Vejamos o TCL de Lindberg-Feller.

3.3.1 TCL variante de Lindberg-Feller

Caso Univariado

Se $x_1, \ldots, x_n$ é uma amostra aleatória de uma distribuição de probabilidade com média μ e variância σ^2 na população:

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

Então,

$$\sqrt{n}(\bar{X}_n - \mu) \overset{d}{\sim} N[0, \sigma^2]$$

em que $(\bar{X}_n - \mu)$ são desvios da média da amostra em relação à média da população.

A média de $\sqrt{n}(\bar{X}_n - \mu)$ é obtida por:

$$\begin{gathered}
E[\sqrt{n}(\bar{X}_n - \mu)] = E[\sqrt{n}\bar{X}] - E[\sqrt{n}\mu] = \sqrt{n}E[\bar{X}] - \sqrt{n}E[\mu]\\
E[\sqrt{n}(\bar{X}_n - \mu)] = \sqrt{n}\mu - \sqrt{n}\mu = 0
\end{gathered} \tag{13}$$

Caso Multivariado

Agora temos $\mathbf{x_1}, \ldots, \mathbf{x_n}$ representando amostras de uma distribuição multivariada. Assim:

$$\sqrt{n}(\bar{\mathbf{X}} - \mu) \overset{d}{\sim} N[\mathbf{0}, \mathbf{Q}]$$

No nosso caso, temos que

$$\frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon = \sqrt{n}[\bar{\mathbf{w}} - E(\bar{\mathbf{w}})]$$

em que:

$$\sqrt{n}[\bar{\mathbf{w}} - E(\bar{\mathbf{w}})] \overset{d}{\sim} N[\mathbf{0}, \mathbf{Q}]$$

Então:

$$\frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon \overset{d}{\sim} N[\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{Q}]$$

Segue-se que:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{\mathbf{X^T X}}{n}^{-1}\right)\frac{\mathbf{X^T}\epsilon}{\sqrt{n}}$$

$$\begin{aligned}
\lim_{n\to+\infty}[\sqrt{n}(\hat{\beta} - \beta)] &= \lim_{n\to+\infty}\left[\left(\frac{\mathbf{X^T X}}{n}^{-1}\right)\frac{\mathbf{X^T}\epsilon}{\sqrt{n}}\right] \\
\lim_{n\to+\infty}[\sqrt{n}(\hat{\beta} - \beta)] &= \mathbf{Q}^{-1}\tfrac{1}{\sqrt{n}}\mathbf{X^T}\epsilon \\
\mathbf{Q}^{-1}\tfrac{1}{\sqrt{n}}\mathbf{X^T}\epsilon &\overset{d}{\sim} N[\mathbf{Q}^{-1}0, \mathbf{Q}^{-1}(\sigma^{\mathbf{2}}\mathbf{Q})\mathbf{Q}^{-1}] \\
\mathbf{Q}^{-1}\tfrac{1}{\sqrt{n}}\mathbf{X^T}\epsilon &\overset{d}{\sim} N[\mathbf{Q}^{-1}0, \sigma^{\mathbf{2}}\mathbf{Q}^{-1}]
\end{aligned} \tag{14}$$

Observação: Lembre-se que o primeiro termo é a média de $\mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon$:

$$E\left[\mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon\right] = \mathbf{Q}^{-1}0$$

e o segundo termo é a variância de $\mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon$:

$$Var\left(\mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\mathbf{X^T}\epsilon\right) = \mathbf{Q}^{-1}(\sigma^{\mathbf{2}}\mathbf{Q})\mathbf{Q}^{-1}$$

Então:

$$\sqrt{n}(\hat{\beta} - \beta) \overset{d}{\sim} N[\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{Q}^{-1}]$$

Vamos agora utilizar um resultado do Greene (2012):

Teorema: Se $\sqrt{n}(\hat{\theta}_n - \theta) \overset{d}{\sim} N[0, V]$, então $\hat{\theta} \sim N\left[\theta, \frac{1}{n}V\right]$, em que a primeira expressão diz que $\sqrt{n}(\hat{\theta}_n - \theta)$ converge em distribuição para uma normal com média zero e variância V e a segunda expressão diz que $\hat{\theta}$ **é assintoticamente normalmente distribuido** com média 0 e variância $\frac{1}{n}V$.

No caso do estimador MQO, se:

$$\sqrt{n}(\hat{\beta} - \beta) \overset{d}{\sim} N[\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{Q^{-1}}]$$

então:

$$\hat{\beta} \sim N\left[\beta, \frac{\sigma^{\mathbf{2}}\mathbf{Q^{-1}}}{n}\right]$$

ou

$$\hat{\beta} \sim N[\beta, \sigma^{\mathbf{2}}(\mathbf{X^T X})^{\mathbf{-1}}]$$

Logo, a normalidade assintótica de $\hat{\beta}$ não necessita da normalidade dos erros! É uma consequência do Teorema Central do Limite!

3.4 Consistência de s^2 e a variância assintótica de $\hat{\beta}$

Para computar as propriedades assintóticas de $\hat{\beta}$ precisamos saber se o estimador da variância, $\hat{\sigma}^{\mathbf{2}}$ é consistente. Aqui só precisamos de consistência!

O problema é a consistência do estimador de $\sigma^{\mathbf{2}}$, $\hat{\sigma}^{\mathbf{2}} = \mathbf{s^2}$, uma vez que $\mathbf{Q^{-1}}$ não é problema. Será que $\mathbf{s^2}$ é consistente?

$$\begin{aligned}
\mathbf{s^2} &= \tfrac{\hat{\epsilon}^{\mathbf{T}}\hat{\epsilon}}{n-k} \\
\mathbf{s^2} &= \tfrac{1}{n-k}\epsilon^{\mathbf{T}}\mathbf{M}\epsilon
\end{aligned} \tag{15}$$

Lembre-se que M é a matriz geradora de resíduos, obtida como se segue:

$$\begin{aligned}
\hat{\epsilon} &= \mathbf{y} - \mathbf{X}\hat{\beta} \\
\hat{\epsilon} &= \mathbf{y} - \mathbf{X}(\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X^Ty} \\
\hat{\epsilon} &= \mathbf{I} - \mathbf{X}(\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X^T})\mathbf{y} \\
\hat{\epsilon} &= \mathbf{My} \\
\hat{\epsilon} &= \mathbf{M}(\mathbf{X}\beta + \epsilon) \\
\hat{\epsilon} &= \mathbf{MX}\beta + \mathbf{M}\epsilon = \mathbf{M}\epsilon \\
\hat{\epsilon}^{\mathbf{T}}\hat{\epsilon} &= \epsilon^{\mathbf{T}}\mathbf{M^TM}\epsilon \\
\hat{\epsilon}^{\mathbf{T}}\hat{\epsilon} &= \epsilon^{\mathbf{T}}\mathbf{M}\epsilon
\end{aligned} \tag{16}$$

Abrindo a expressão $\mathbf{s^2}$ e multiplicando, temos:

$$\begin{aligned}
\mathbf{s^2} &= \tfrac{1}{n-k}\epsilon^{\mathbf{T}}[\mathbf{I} - \mathbf{X}(\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X^T}]\epsilon \\
\mathbf{s^2} &= \tfrac{1}{n-k}[\epsilon^{\mathbf{T}}\epsilon - \epsilon^{\mathbf{T}}\mathbf{X}(\mathbf{X^TX})^{-\mathbf{1}}\mathbf{X^T}\epsilon]
\end{aligned} \tag{17}$$

Multiplicando por $\frac{n}{n}$, temos:

$$\begin{aligned}
\mathbf{s^2} &= \tfrac{n}{n-k}[\tfrac{\epsilon^{\mathbf{T}}\epsilon}{n} - \tfrac{\epsilon^{\mathbf{T}}\mathbf{X}}{n}\left(\tfrac{\mathbf{X^TX^{-1}}}{n}\right)\tfrac{\mathbf{X^T}\epsilon}{n}] \\
\lim_{n\to+\infty}\mathbf{s^2} &= \lim_{n\to+\infty}\tfrac{n}{n-k}[\tfrac{\epsilon^{\mathbf{T}}\epsilon}{n} - \tfrac{\epsilon^{\mathbf{T}}\mathbf{X}}{n}\left(\tfrac{\mathbf{X^TX^{-1}}}{n}\right)\tfrac{\mathbf{X^T}\epsilon}{n}]
\end{aligned} \tag{18}$$

Sabe-se que $\frac{n}{n-k}$ converge para 1, $\frac{\epsilon^{\mathbf{T}}\mathbf{X}}{n}$ e $\frac{\mathbf{X}^{\mathbf{T}}\epsilon}{n}$ convergem para zero e $\frac{\mathbf{X}^{\mathbf{T}}\mathbf{X}}{n}$ converge para $\mathbf{Q^{-1}}$. Portanto, resta saber qual o limite de $\frac{\epsilon^{\mathbf{T}}\epsilon}{n}$, ou seja:

$$\lim_{n\to+\infty} \mathbf{s^2} = \lim_{n\to+\infty} \frac{\epsilon^{\mathbf{T}}\epsilon}{n}$$

Para chegar a este resultado, utilizamos o seguinte teorema:

Teorema de Kinshine: A média de uma amostra aleatória de observações independentes e identificamente distribuídas (*i.i.d.*) é um estimador consistente da média da população.

Então, precisamos apenas assumir que os erros são *i.i.d.*, não precisamos assumir que eles são normais.

Neste caso, basta olharmos $\epsilon_1^2, \epsilon_2^2, \ldots, \epsilon_n^2$ como uma amostra aleatória *i.i.d.* com média na população igual σ^2.

Observação: Isso é válido porque:

$$E[\epsilon_i^2] = \sigma^2$$

A média de ϵ_i^2 nossa amostra é dada por:

$$\bar{\epsilon}^2 = \frac{\sum_{i=1}^{n} \epsilon_i^2}{n} = \frac{\epsilon^{\mathbf{T}}\epsilon}{n}$$

Temos então que $\frac{\epsilon^{\mathbf{T}}\epsilon}{n}$ é um **estimador consistente** da média da população σ^2 (pelo Teorema de Kinshine).

Logo:

$$plim\left(\frac{\epsilon^{\mathbf{T}}\epsilon}{n}\right) = \sigma^2$$

Então:

$$plim(s^2) = \sigma^2$$

Observação:

$$\hat{\beta} \sim N\left[\beta, \sigma^2 \frac{\mathbf{Q}^{-1}}{n}\right]$$

ou

$$\hat{\beta} \sim N[\beta, \sigma^2 \mathbf{X}^{\mathbf{T}}\mathbf{X}^{-1}]$$

A variância assintótica de $\hat{\beta}$ é:

$$AsyVar[\hat{\beta}] = s^2(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$$

exatamente igual à da amostra finita.

3.5 Eficiência assintótica de $\hat{\beta}$

Definição: "um estimador é assintoticamente eficiente quando ele é **consistente**, **assintoticamente normalmente distribuído** e tem uma **matriz de variância-covariância assintótica que não é maior que nenhuma outra matriz de variância-covariância assintótica de um outro estimador** nas mesmas condições, isto é, **consistente e normalmente assintoticamente distribuído**".

Já vimos que o vetor $\hat{\beta}$ dos MQO é consistente, normalmente distribuído. Além dessas propriedades assintóticas, é provado que **a matriz da variância-covariância assintótica é mínima**.

4 Prática de Teoria Assintótica

4.1 Simulação de Monte Carlo Simples e Propriedades Assintóticas

Nesse exercício trabalharemos com duas variáveis aleatórias, x e z, com mesma média e variâncias diferentes.

1. Crie erros aleatórios com distribuição normal, com média zero e variância unitária. Peça as médias para as variáveis x e z e para os erros criados para diferentes números de observações (por exemplo, $10, 100, 1000$) e verifique que as variâncias das variâveis diminuem à medida que aumenta o número de observações.

2. Crie novas variáveis (variáveis dependentes) y como função de x e z, de cada vez, da seguinte forma: $y = 20 + 0,6x + \epsilon$ e $y = 20 + 0,6z + \epsilon$.

3. Estime os modelos de regressão linear para números diferentes de observações $(10, 10, 1000)$, tanto para $y = f(x, \epsilon)$ quanto para $y = f(z, \epsilon)$

4. Crie uma variável w correlacionada com os erros aleatórios, da seguinte maneira: $w = 150 + \epsilon$

5. Transforme as variáveis x e z de modo a serem correlacionadas com os erros, da seguinte maneira: $wx = x + 500\epsilon$ e $wz = z + 5\epsilon$

6. Calcule as médias e teste as correlações entre todas as variáveis e os erros. O que se observa?

7. Estime modelos de regressão linear com os y criados e as variáveis x, z, w, wx e wz para diferentes tamanhos de amostras $(10, 100, 1000)$. Observe a convergência dos parâmetros para cada modelo e discuta com detalhes, segundo o que você conhece sobre a propriedade de consistência do estimador dos MQO.

Observações Gerais:

A convergência do estimador dos MQO pode ser compreendida da seguinte forma:

$$\begin{gathered}
\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^Y} \\
\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X}^(\mathbf{X}\beta + \epsilon) \\
\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^X}\beta + (\mathbf{X^T X})^{-1}\mathbf{X^T}\epsilon \\
\hat{\beta} = \beta + (\mathbf{X^T X})^{-1}\mathbf{X^T}\epsilon \\
plim\hat{\beta} = \beta + plim\left(\frac{\mathbf{X^T X}}{n}\right)^{-1}\frac{\mathbf{X^T}\epsilon}{n}
\end{gathered} \tag{19}$$

Atenção: quanto maior a variabilidade na matrix $\mathbf{X}$, mais rápida é a convergência.

Para casa: crie uma variável xk muito correlacionada com os erros e repita o procedimento. Conclua.

Referências

Goldfeld, S. M. and R. E. Quandat (1972). Nonlinear methods in econometrics.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Pearson Education India.

Leamer, E. E. (1988). 3 things that bother me. *Economic Record 64*(4), 331–335.