# Introdução à Teoria Assintótica para modelos para análise de dados longitudinais

Prof. Caio Azevedo

# Introdução

- Convergência em distribuição (estimadores/estatísticas do teste), convergência em probabilidade/quase certa (estimadores).
- Existem algumas formas de deduzir resultados assintóticos (slide seguinte):

# Introdução

- Resultados específicos para modelos de regressão (geralmente flexibilizando as suposições usuais, normalidade, heterocedasticidade, dependência) relativos à determinados métodos de estimação: mínimos quadrados, máxima verossimilhança, máxima verossimilhança marginal.
- Estudando o comportamento de classes específicas de estimadores:mínimos quadrados, máxima verossimilhança, máxima verossimilhança marginal, verificando se as condições que garantem as propriedades para tais estimadores, valem para os modelos em estudo.

# Comentários

- Convergência em distribuição - Teorema central do limite.

  Essencialmente três tipos:

  - Variáveis aleatórias IID.
  - Variáveis aleatórias independentes mas não identicamente distribuídas (o caso).
  - Variáveis que apresentam dependência do tipo martingal.

- Convergência em probabilidade - Desigualdades (Chebyshev)

# Modelo normal linear multivariado (marginal)

$$\boldsymbol{Y}_{j(k_j \times 1)} = \boldsymbol{X}_{j(k_j \times p)}\boldsymbol{\beta}_{(p \times 1)} + \boldsymbol{\xi}_{j(k_j \times 1)}, j = 1, ..., n \tag{1}$$

(indivíduo)

- $\boldsymbol{Y}_j = (Y_{1j}, ..., Y_{k_j j})'$, $k_j$ : número de condições de avaliação em que o indivíduo $j$ é avaliado.
- $\boldsymbol{X}_j$ : matriz de planejamento associada aos efeitos fixos (parâmetros de regressão) para o indivíduo $j$ (não-aleatória e conhecida).
- $\boldsymbol{\beta}$ : vetor de efeitos fixos ou parâmetros de regressão (não-aleatório e desconhecido).
- $\boldsymbol{\xi}_j$ : vetor de erros associado ao indivíduo $j$, $\boldsymbol{\xi}_j \overset{ind.}{\sim} N_{k_j}(\mathbf{0}, \boldsymbol{\Sigma}_j)$.

# Modelo Multivariado Marginal (MMM)

- Essencialmente correspondem à extensões dos modelos de regressão lineares normais homocedásticos (MRLNH) (veja `http://www.ime.unicamp.br/~cnaber/Material_REG_POS_2S_2014.htm`), no sentido de considerarem heterocedasticidade e dependência entre as observações.
- Assim, os resultados assintóticos desenvolvidos para o MRLNH, com o relaxamento das suposições de homocedasticidade e dependência, podem ser utilizados para demonstrar as propriedades assintóticas dos estimadores de máxima verossimilhança dos MMM.

# Modelo Multivariado Marginal (MMM)

- Além disso note que, usualmente, considera-se a parametrização adotada pela função gls do pacote R, veja `http://www.ime.unicamp.br/~cnaber/aula_Mod_Mult_Marg_ADL_2S_2018.pdf`. Assim, a dedução dos resultados assintóticos torna-se menos complexa.

# Ausência de normalidade

- Considere o modelo de regressão linear geral:

$$\boldsymbol{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\xi}_n$$

- Suponha que $\mathcal{E}(\boldsymbol{\xi}_n) = \mathbf{0}_n$ e $Cov(\boldsymbol{\xi}_n) = \sigma^2\boldsymbol{I}_n$.
- Suposições adicionais
  1. $max_{1\leq k\leq n}\boldsymbol{x}'_{n_k}(\boldsymbol{X}'_n\boldsymbol{X}_n)^{-1}\boldsymbol{x}_{n_k} \to 0$, quando $n \to \infty$.
  2. $\lim_{n\to} n^{-1}(\boldsymbol{X}'_n\boldsymbol{X}_n) = \boldsymbol{V}$.
- Pelo Teorema Central do Limte (TCL) de Linderberg-Feller (se $n \to \infty$), então

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \sigma^2\boldsymbol{V}^{-1})$$

# Comentário

- Para $n$ suficientemente grande

$$\widehat{\beta} \approx N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}_n\boldsymbol{X}_n)^{-1})$$

- Em relação à primeira suposição adicional, considere que

$$\boldsymbol{X}_n = \begin{pmatrix} \boldsymbol{x}'_{n_1} \\ \boldsymbol{x}'_{n_2} \\ \vdots \\ \boldsymbol{x}'_{n_n} \end{pmatrix} ; \boldsymbol{X}_n = \begin{pmatrix} \boldsymbol{1}_{n_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{n_2} \end{pmatrix} ; n^{-1}\boldsymbol{X}'_n\boldsymbol{X}_n = \begin{pmatrix} n_1/n & 0 \\ 0 & n_2/n \end{pmatrix}$$

- Assim

$$\boldsymbol{x}'_{n_1}(\boldsymbol{X}'_n\boldsymbol{X}_n)^{-1}\boldsymbol{x}_{n_1} = n_1^{-1}; \boldsymbol{x}'_{n_2}(\boldsymbol{X}'_n\boldsymbol{X}_n)^{-1}\boldsymbol{x}_{n_2} = n_2^{-1}$$

# Ausência de homocedasticiade e presença de correlação

- Suponha a mesma estrutura anterior mas $Cov(\boldsymbol{\xi}) = \sigma^2\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}$ conhecida.
- Defina

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{\delta} \tag{2}$$

em que

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}(\text{decomposição de Cholesky}) \\
\boldsymbol{Z} &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} \\
\boldsymbol{W} &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X} \\
\boldsymbol{\delta} &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi}
\end{aligned}$$

# Cont.

- Assim, $\mathcal{E}(\boldsymbol{\delta}) = \boldsymbol{\Sigma}^{-1/2}\mathcal{E}(\boldsymbol{\xi}) = \mathbf{0}$ e $Cov(\boldsymbol{\delta}) = \sigma^2\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \sigma^2\boldsymbol{I}$.
- Logo, o estimador de MQO de $\boldsymbol{\beta}$, baseado no modelo (2), é dado por:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{Z} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

- Se $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \sigma^2\boldsymbol{\Sigma})$, então

$$\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1})$$

# Cont.

- Se a suposição de normalidade for desconsiderada mas as suposições relativas ao TCL anterior se verificarem, então

$$\widehat{\boldsymbol{\beta}} \approx N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1})$$

- Quando $\boldsymbol{\Sigma}$ é diagonal mas ($\neq \boldsymbol{I}$), dizemos que $\widehat{\boldsymbol{\beta}}$ é o estimador de mínimos quadrados ponderados (MQP) de $\boldsymbol{\beta}$ (baseado no modelo original).
- Quando $\boldsymbol{\Sigma}$ não é diagonal, dizemos que $\widehat{\boldsymbol{\beta}}$ é o estimador de mínimos quadrados generalizados (MQG) de $\boldsymbol{\beta}$ (baseado no modelo original).

# $\boldsymbol{\Sigma}$ desconhecido

- Considere novamente o modelo original, $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\xi}$, com $\mathcal{E}(\boldsymbol{\xi}) = \boldsymbol{0}$ e $Cov(\boldsymbol{\xi}) = \boldsymbol{\Sigma}$.
- Uma alternativa consite em considerar o estimador de dois estágio de Aitken, ou seja

$$\widehat{\beta} = (\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y}$$

em que $\widehat{\boldsymbol{\Sigma}}$ é um estimador consistente de $\boldsymbol{\Sigma}$, ($\widehat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}$).

# Cont.

- Assim, temos que (se válidas as condições do TCL apresentado anteriomente), que:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{V}^{-1})$$

em que $\boldsymbol{V} = \lim_{n\to\infty} n^{-1}(\boldsymbol{X}_n'\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}_n)$

- Em geral, toma-se $\widehat{\boldsymbol{V}} = \boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}$.

# Modelo normal linear misto

$$\boldsymbol{Y}_{j(k_j \times 1)} = \boldsymbol{X}_{j(k_j \times p)}\boldsymbol{\beta}_{(p\times 1)} + \boldsymbol{Z}_{j(k_j \times q)}\boldsymbol{b}_{j(q\times 1)} + \boldsymbol{\xi}_{j(k_j \times 1)}, j = 1, ..., n$$

(indivíduo)

- $\boldsymbol{Y}_j = (Y_{j1}, ..., Y_{jk_j})'$, $k_j$ : número de condições de avaliação em que o indivíduo $j$ é avaliado.
- $\boldsymbol{X}_j$ : matriz de planejamento associada aos efeitos fixos para o indivíduo $j$ (não-aleatória e conhecida).
- $\boldsymbol{Z}_j$ : matriz de planejamento associada aos efeitos aeatórios para o indivíduo $j$ (não-aleatória e conhecida).
- $\boldsymbol{\beta}$ : vetor de efeitos fixos (não-aleatório e desconhecido).
- $\boldsymbol{b}_j$ : vetor de efeitos aleatórios associado ao indivíduo $j$ (aleatório e desconhecido), $\boldsymbol{b}_j \overset{ind.}{\sim} N(\mathbf{0}, \boldsymbol{\Psi})$.
- $\boldsymbol{\xi}_j$ : vetor de erros associado ao indivíduo $j$, $\boldsymbol{\xi}_j \overset{ind.}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_j)$, $\boldsymbol{b}_j \perp \boldsymbol{\xi}_j, \forall i$.

# Algumas propriedades do modelo

- $\mathcal{E}(\boldsymbol{Y}_j|\boldsymbol{b}_j) = \boldsymbol{X}_j\beta + \boldsymbol{Z}_j\boldsymbol{b}_j$.
- $\mathcal{E}(\boldsymbol{Y}_j) = \boldsymbol{X}_j\beta$.
- $Cov(\boldsymbol{Y}_j|\boldsymbol{b}_j) = \boldsymbol{\Sigma}_j$.
- $Cov(\boldsymbol{Y}_j) = \boldsymbol{V}_j = \boldsymbol{Z}_j\boldsymbol{\Psi}\boldsymbol{Z}_j' + \boldsymbol{\Sigma}_j$.
- $\boldsymbol{Y}_j|\boldsymbol{b}_j \sim N_{k_j}\left(\boldsymbol{X}_j\beta + \boldsymbol{Z}_j\boldsymbol{b}_j, \boldsymbol{\Sigma}_j\right)$. Além disso, como

$$\begin{aligned} \boldsymbol{Y}_j|\boldsymbol{b}_j &\sim N_{k_j}(\boldsymbol{X}_j\beta + \boldsymbol{Z}_j\boldsymbol{b}_j, \boldsymbol{\Sigma}_j) \\ \boldsymbol{b}_j &\sim N_q(\boldsymbol{0}, \boldsymbol{\Psi}) \end{aligned}$$

portanto

$$\boldsymbol{Y}_j \sim N_{k_j}(\boldsymbol{X}_j\beta, \boldsymbol{Z}_j\boldsymbol{\Psi}\boldsymbol{Z}_j' + \boldsymbol{\Sigma}_j) \quad (3)$$

# Propriedades assintóticas

- Note que a representação (3) parece com os MMM, com duas matrizes de covariâncias.
- Assim, os resultados discutidos anteriormente, para os MMM, podem ser adaptados para este caso.
- Note também que, usualmente, utilaza-se a parametrização considerada pela função lme do pacote R (`http://www.ime.unicamp.br/~cnaber/aula_Mod_Lin_Misto_ADL_2S_2018.pdf`), que corresponde à uma representação mais simples do modelo.
- Consequentemente, os resultados assintóticos são dedutíveis de forma mais simples.

# Estimadores de máxima verossimilhança/máxima verossimilhança marginal

- Livros sobre teoria assintótica (geral) apresentam condições para convergência em distribuição (normalidade multivariada) para estimadores de máxima verossimilhança para variáveis IID.
- A idéia para se demonstrar os resultados, para o caso de modelos de regressão (MMM, MLM) consiste em visualizar as funções (vetor) escore como variáveis (vetores) aleatórios IID com vetor de médias zero e matriz de covariâncias igual a inversa da Informação de Fisher, com condições adicionais.

## Estimadores de máxima verossimilhança/máxima verossimilhança marginal

- Pode-se usar raciocínio análogo, no caso de modelos lineares mistos, em que os vetores escore são vetores aleatórios independentes mas não identicamente distribuídas (INID), de média zero e matriz de covariâncias igual à inversa da Informação de Fisher, com condições adicionais.
- Neste caso, deve-se verificar condições mais específicas (para variáveis INID) como as de Liapunov e Lindenberg-Feller.

# Referências

- Demonstrações acerca de resultados assintóticos para modelos de regressão e estimadores de máxima verossimilhança (Sen and Singer (2019), Sen et al (2009)).
- Demonstrações acerca de estimadores, estatísticas do teste para modelos lineares mistos, problemas em aberto (Demidenko (2004), Demidenko (2013), Jiang (2017)).
- Outras referências sobre Teoria Assintótica para modelos mistos Singer et al 2018 (`https://www.ime.usp.br/~jmsinger/MAE0610/Singer&Nobre&Rocha2018jun.pdf`).

# Referências


- Demidenko, E. (2004). Mixed models: theory and applications, Hoboken, NJ: Wiley-Interscience.
- Demidenko, E. (2013). Mixed Models: Theory and Applications with R, Wiley-Series.
- Jiang, J. (2017). Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems, Chapman and Hall/CRC.
- Sen, P. K., Singer, J. M. (2019). Large Sample Methods in Statistics (1994): An Introduction with Applications, CRC Press Revivals.
- Sen, P. K., Singer, J. M., Lima, A. C. P. (2009). From Finite Sample to Asymptotic Methods in Statistics, Cambridge University Press.