Considerações sobre a inclusão de efeitos aleatórios em modelos de transição de Markov

Idemauro Antonio Rodrigues de Lara[1] **Clarice Garcia Borges Demétrio**[2]

Sílvia Emiko Shimakura[3]

[1]Departamento de Estatística, UFPR; [2]Departamento de Ciências Exatas, ESALQ-USP

[3]Laboratório de Estatística e Geoinformação, DEST-UFPR

idemauro@ufpr.br

1. Introdução

MODELO DE TRANSIÇÃO DE MARKOV ⇒ MLG para a distribuição condicional de Y_{it} dado $\boldsymbol{h}_{it}$ e $\boldsymbol{x}_{it}$ ⇒ modelos baseados em processos estocásticos (DIGGLE et al., 2002).

- $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$ (variáveis respostas);
- $\boldsymbol{x}_{it} = (x_{it1}, \ldots, x_{itp})'$ (covariáveis);
- $\boldsymbol{h}_{it} = (y_{i1}, y_{i2}, \ldots, y_{i,(t-1)})'$ (respostas prévias).

A média e a variância condicional satisfazem as equações:

$$g(\mu_{it}^C) = \eta_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \sum_{r=1}^{s} f_r(\boldsymbol{h}_{it}; \boldsymbol{\alpha}) \text{ e } v_{it}^C = v(\mu_{it}^C)\phi, \tag{1}$$

Parâmetros de interesse (a priori)⇒ $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ ⇒ modelos de efeitos fixos.

- Com Variáveis categorizadas ⇒ especificar probabilidades de transição.
- Para dados binários, assumindo ligação canônica e dependência estocástica 1, tem-se:

$$\boldsymbol{P}(t) = \begin{pmatrix} \pi_{00}(t) = \dfrac{1}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})} & \pi_{01}(t) = \dfrac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})} \\ \pi_{10}(t) = \dfrac{1}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha)} & \pi_{11}(t) = \dfrac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha)}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha)} \end{pmatrix}.$$

1.1 Considerações iniciais

- não há razões para se restringir apenas aos efeitos fixos na estrutura do preditor linear (1) [Korn e Whittemore (1979); Stiratelli, Laird e Ware (1984); Lara (2007)];
- a estrutura de medidas repetidas com variáveis binárias aliada ao número de ocasiões pode esconder a existência de efeitos aleatórios;
- necessidade de desenvolverem estudos com simulação.

2. Métodos

2.1 Definição do modelo de transição misto

Para uma cadeia de ordem q, tem-se como preditor linear:

$$\eta_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \sum_{r=1}^{s} f_r(\boldsymbol{h}_{it}; \boldsymbol{\alpha}) + \boldsymbol{z}_{it}'\boldsymbol{d}_i, \tag{2}$$

em que $\boldsymbol{\beta}, \boldsymbol{\alpha}$ são os parâmetros associados aos efeitos fixos, $\boldsymbol{z}_{it}$ é a i-ésima linha da matriz do modelo associada ao vetor de efeitos aleatórios, $\boldsymbol{d}_i(k \times 1)$. Sendo $\boldsymbol{d}_i \sim N_k(\boldsymbol{0}, \boldsymbol{G})$.
Para simplificar a notação (2), considere $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$:

$$\eta_{it} = \boldsymbol{x}_{it}^{*'}\boldsymbol{\delta} + \boldsymbol{z}_{it}'\boldsymbol{d}_i.$$

A função de verossimilhança para $\boldsymbol{\delta}$ e $\boldsymbol{G}$ no caso estacionário é proporcional a:

$$L(\boldsymbol{\delta}, \boldsymbol{G}; \boldsymbol{y}) \propto \prod_{i=1}^{N} \int \prod_{t=2}^{n_i} [\mu_{it}(\boldsymbol{\delta}, \boldsymbol{G})]^{y_{it}} [1 - \mu_{it}(\boldsymbol{\delta}, \boldsymbol{G})]^{1-y_{it}} f(\boldsymbol{d}_i; \boldsymbol{G}) d(\boldsymbol{d}_i), \tag{3}$$

em que $\mu_{it}(\boldsymbol{\delta}, \boldsymbol{d}_i) = \mathsf{E}(Y_{it} \mid \boldsymbol{d}_i; \boldsymbol{\delta})$.
Focalizando a atenção sob função de ligação canônica e supondo distribuição normal para o efeito aleatório, $\boldsymbol{d}_i$, a expressão (3), reduz-se a:

$$\prod_{i=1}^{N} \int \exp\left[\boldsymbol{\delta}' \sum_{t=2}^{n_i} \boldsymbol{x}_{it}^{*} y_{it} + \boldsymbol{d}_i' \sum_{t=2}^{n_i} \boldsymbol{z}_{it} y_{it} - \sum_{t=2}^{n_i} \ln\{1 + \exp(\boldsymbol{x}_{it}^{*'}\boldsymbol{\delta} + \boldsymbol{z}_{it}'\boldsymbol{d}_i)\}\right] (2\pi)^{-1} |\boldsymbol{G}|^{-1/2} \exp\left(\frac{-\boldsymbol{d}_i'\boldsymbol{G}^{-1}\boldsymbol{d}_i}{2}\right) d(\boldsymbol{d}_i), \tag{4}$$

em que $\boldsymbol{G}$ é matriz de covariâncias dos efeitos aleatórios, $\boldsymbol{d}_i$.

Nesse trabalho, para maximização da função (4) optou-se pela técnica baseada na aproximação dos integrandos (Método de Laplace), devido à otimização do custo operacional das simulações.

2.2 Estudo por simulação

A implementação computacional foi feita no *software* R.

- Modelo para simulação:

$$\eta_{it} = (\beta_0 + d_{i0}) + \beta x_{it} + \alpha y_{i(t-1)} \tag{5}$$

- assumindo estacionariedade e alcance 1 para a cadeia;
- vetor de parâmetros: $\boldsymbol{\delta} = (-2,5; 1,0; 2,0)$ e $d_i \sim N(0; 0,5)$;
- $n = 500$ indivíduos;
- $t = 4, 5, 8$ e 12 ocasiões;
- 10.000 simulações para cada situação de ocasião;
- ajustaram-se modelos de transição com interceptos aleatórios (estrutura real ⇒ equação 5) pela função *glmmML* e ignorando-se essa estrutura pela função *geeglm* (modelos somente com efeitos fixos);
- teste da razão de verossimilhanças e o critério da informação de Akaike (AIC) foram usados para discriminar entre esses dois modelos.

3. Resultados

Tabela 1: *Erro quadrático médio das estimativas dos parâmetros com base nos dados simulados, considerando modelos somente com efeitos fixos (Modelo 1) e mistos (Modelo 2) e 4 diferentes números de possibilidades de ocasiões.*

	$t=4$	
Parâmetros	Modelo 1	Modelo 2
β_0	0,01896	0,02364
β_1	0,02749	0,02905
α	0,05632	0,04858
	$t=5$	
Parâmetros	Modelo 1	Modelo 2
β_0	0,01515	0,01583
β_1	0,02311	0,02221
α	0,05317	0,03952
	$t=8$	
Parâmetros	Modelo 1	Modelo 2
β_0	0,01070	0,00847
β_1	0,01783	0,01339
α	0,05040	0,02031
	$t=12$	
Parâmetros	Modelo 1	Modelo 2
β_0	0,00825	0,00533
β_1	0,01573	0,00913
α	0,04926	0,01011

Tabela 2: *Percentual do número de vezes em que o AIC do Modelo 1 foi maior do que o do Modelo 2 de acordo com o número de ocasiões.*

	Número de ocasiões			
	4	5	8	12
Percentual	$12,29\%$	$18,66\%$	$62,59\%$	$96,93\%$

Verificou-se uma tendência dos modelos em subestimar o efeito da covariável X e superestimar o efeito de $Y_{(t-1)}$, sendo esses vícios maiores para o Modelo 1, ou seja, ignorar a existência de efeito aleatório fatalmente acarretará em estimativas viciadas, que conseqüentemente vão distorcer as probabilidades de transição.

4. Considerações Finais

- Modelos de transição misto tem duas fontes de correlação: história e heterogeneidade entre os coeficientes de regressão;
- Deve-se avaliar a necessidade da inclusão de efeitos aleatórios ou ainda testar a existência desses efeitos no modelo;
- Matriz de probabilidades de transição pode ser interpretada individualmente, ou seja, uma matriz para cada indivíduo.

Referências Bibliográficas

DIGGLE, P.J.; HEAGERTY, P.J.; LIANG, K.Y.; ZEGER, S.L. **Analysis of longitudinal data**. New York: Oxford University Press, 2002, 379 p.

KORN, E.L.; WHITTEMORE, A.S. Methods for analysing panel studies of acute health effects of air pllution. **Biometrics**, Washington, v. 35, p. $715-802$, 1979.

LARA, I.A.R. **Modelos de transição para dados binários**. 2007, 111 p. Tese (Doutorado em Estatística e Experimentação Agronômica), Escola Superior de Agricultura Luiz de Queiroz-USP, Piracicaba, 2007.

STIRATELLI, R.; LAIRD, N.; WARE, J.H. Random effects-models for serial observations with binary response. **Biometrics**, Washington, v. 40, p. $961-971$, 1984.