Algoritmo EM condicional via dados aumentados aplicado à Teoria de Resposta ao Item - TRI

Prof. Caio Azevedo

Resumo

- Breve Introdução à TRI e à estimação na TRI.
- Motivação.
- Apresentação do método de estimação (CADEM).
- Estudo de simulação.
- Conclusões e comentários.

Introdução, Motivação e Objetivos

- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e items de um instrumento de medida (prova, questionário).
- Tal modelagem consiste na probabilidade de obter um certo escore em cada item.
- Existe um grande número de classes de MRI : dicotômicos e policotômicos, um e múltiplos grupos, multidimensionais, longitudinais multiníveis, dentre outros. Um MRI apresenta um grande número de parâmetros.

Introdução, Motivação e Objetivos

- As classes mais simples de MRI possuem apenas traços latentes e parâmetros de dificuldade (Modelos de 1 parâmetro ou de Rasch).
- Se o número de indivíduos aumenta $\rightarrow$ número de traços latentes aumenta. Se o número de items aumenta $\rightarrow$ número de parâmetros dos itens aumenta.

Histórico e Características da estimação na TRI

- Características.
  - Número elevados de parâmetros para estimar.
  - A verificação das propriedades dos estimadores é muito complicada.
  - Falta de identificabilidade.
  - Necessidade de utilização de métodos numéricos.
  - Espaços paramétricos restritos.

Histórico da estimação na TRI

- Máxima Verossimilhança Marginal, MVM, (Bock and Lieberman, 1970).
- MVM via pseudo algoritmo EM (Bock and Aitkin, 1981).
- MCMC via dados aumentados (Albert, 1992).
- MCMC via algoritmo de MH (Patz and Junker, 1999).
- Algoritmo EM via expansão paramétrica (Rubin and Thomas, 2000).
- Algoritmo EM estocástico (SEM), (Fox, 2000).
- Algoritmo de Metropolis-Hastings Robbins-Monro (MHRM), (Cai, 2010).
- Algoritmo CADEM (Conditional Augmented Data EM algorithm), (Azevedo and Andrade, 2013).

Motivação

- Todos os métodos anteriores possuem vantagens e desvantagens.
- MVM/MMAP via pseudo EM : baixo tempo computacional, limitação com relação à dimensão latente, distribuição latente, estruturas multiníveis.
- MCMC : elevado tempo computacional, flexibilidade virtualmente infinita.
- Algoritmo EM estocástico : tempo computacional não tão elevado mas é necessário obter convergência em distribuição.

Motivação

- Proposta CADEM : Substituir o passo S (de simulação) no algoritmo SEM por cáculo de esperanças (condicionais) de modo a obter convergência pontual.
- Objetivo : desenvolver um algoritmo EM (frequentista - bayesiano) para o ajuste de MRI e exemplificar seu uso através de um modelo de Rasch.

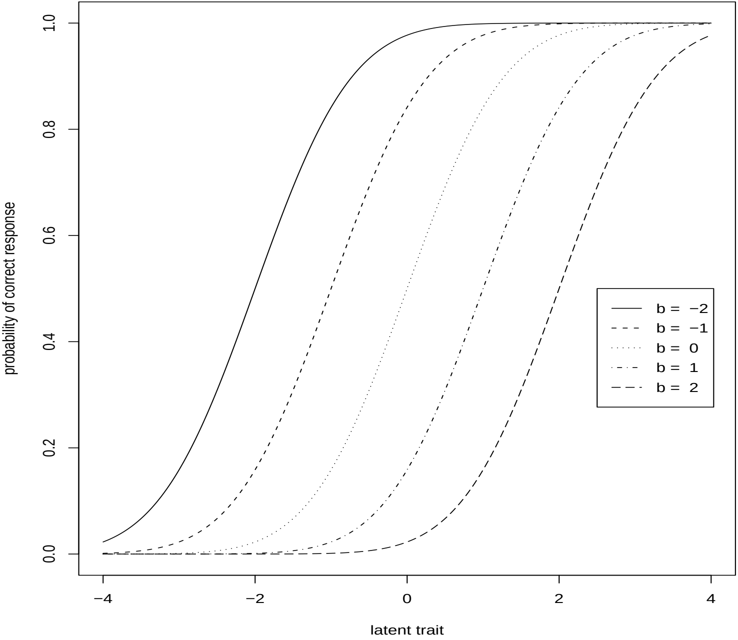
Modelo probito de 1 parâmetro

$$P(Y_{ij} = 1|(\theta_j, b_i)) \quad = \quad \Phi(\theta_j - b_i)$$

$i = 1, \ldots, I \, (item)$, $j = 1, \ldots, n \, (individuo)$,

- Y_{ij} : é a resposta do indivíduo j ao item i. É igual a 1 se o indivíduo responde corretamente e 0 caso scontrário.
- θ_j : é o traço latente (conhecimento, nível de depressão, etc) do indivíduo j.
- b_i : é o parâmetro de dificuldade associado ao item i.
- Na literatura, comumente, assume-se que $\theta_j \sim N(0, 1)$. No entanto, na prática essa suposição pode não ser realista.

Curves of Rasch model



- A abordagem de MVM/MMAP assume uma distribuição normal padrão simétrica (alternativas podem ser econtradas na literatura) no modelo de Rasch.
- A utilização do esquema de dados aumentados de Albert (1992) torna a verossimilhança mais tratável.
- Esquema de dados aumentados

$$Z_{ij}|(\theta_j, b_i) = \begin{cases} N(\theta_j - b_i, 1)\boldsymbol{I}_{(z_{ij} \geq 0)}, & \text{if } Y_{ij} = 1, \\ N(\theta_j - b_i, 1)\boldsymbol{I}_{(z_{ij} < 0)}, & \text{if } Y_{ij} = 0. \end{cases}$$

- Para um teste de I itens aplicado à n respondentes, sob as usuais suposições de independência condicional, a distribuição a posteriori via dados aumentados é dada por

$$
\begin{aligned}
\ln p(\boldsymbol{z}_{..}, \boldsymbol{\theta}, \boldsymbol{b} | \boldsymbol{y}_{...}) &= -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n}\left[z_{ij}^2 - 2z_{ij}\theta_j + \theta_j^2 + 2(z_{ij}b_i - \theta_j b_i)\right. \\
&\left. + b_i^2\right] - \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 + \sum_{i=1}^{I}\ln p(b_i)\,.
\end{aligned}
$$

Algoritmo EM genuíno

- A versão bayesiana usual do algoritmo EM consiste em dois passos
- Passo E : Calcular a esperança das variáveis não-observadas, condicionada nas variáveis observadas e estimativas atualizadas dos parâmetros, no log da posteriori (próximo slide)

$$
\begin{aligned}
&\boldsymbol{E}\left\{\ln p(\boldsymbol{z}_{..},\boldsymbol{\theta},\boldsymbol{b},|\boldsymbol{y}_{...})|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\} = \\
&-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n}\left\{\boldsymbol{E}\left\{z_{ij}^2|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\} - 2\boldsymbol{E}\left\{z_{ij}\theta_j|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\}\right. \\
&+\boldsymbol{E}\left\{\theta_j^2|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\} + 2\boldsymbol{E}\left\{z_{ij}|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\} b_i \\
&\left.-\boldsymbol{E}\left\{\theta_j|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\} b_i + b_i^2\right\} \\
&-\frac{1}{2}\sum_{j=1}^{n}\boldsymbol{E}\left\{\theta_j^2|\left(\boldsymbol{y}_{..},\boldsymbol{b}^{(t)}\right)\right\} + \sum_{i=1}^{I}\ln p(b_i)\,.
\end{aligned}
$$

- Passo M : Maximizar a esperança da log-posteriori com relação à $\boldsymbol{b}^t$

- Propomos a seguinte modificação:
- Passo E : Calcular a esperança de uma partição das variáveis não-observadas, condicionada nas variáveis observadas, nas estimativas atualizadas dos parâmetros e nas outras partições das variáveis não-observadas, no log da posteriori

$$\boldsymbol{E}[\ln p(\boldsymbol{z}_{..},\boldsymbol{\theta},\boldsymbol{b}|\boldsymbol{y}_{...})|\boldsymbol{\vartheta}] = -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n}\Big[\widehat{Z}_{ij}^2 - 2\widehat{Z}_{ij}\widehat{\theta}_j + \widehat{Z}_j^2 + 2(\widehat{Z}_{ij}b_i + \widehat{\theta}_j b_i) + b_i^2\Big] - \frac{1}{2}\sum_{j=1}^{n}\widehat{\theta}_j^2 + \sum_{i=1}^{I}\ln p(b_i)\,. \quad (1)$$

- Passo M : Maximizar a esperança (condicional) da log-posteriori com relação à $\boldsymbol{b}^t$

- Passo E
  - EM usual

$$\boldsymbol{E}\left[h(Z_{ij})|\zeta_i, y_{ij}\right] \text{ e } \boldsymbol{E}\left[g(\theta_{ij})|\zeta_i, y_{ij}\right]$$

  - CADEM

$$\widehat{h(Z_{ij})} = \boldsymbol{E}\left[Z_{ij}|\zeta_i, y_{ij}, \theta_j\right] \text{ e } \widehat{g(\theta_{ij})} = \boldsymbol{E}\left[\theta_{ij}|\zeta_i, y_{ij}, z_{ij}\right]$$

- Passo M
  - EM usual : maximizar a esperança marginal da log-posteriori.
  - CADEM : maximizar a esperança condicional da log-posteriori.

CADEM : Passo E

Passo E1 Dada as estimativas atualizadas $\boldsymbol{b}^{(t)}$ e valores esperados atualizados $\boldsymbol{\theta}^{(t)}$ calcular

$$\widehat{Z}_{ij} = \boldsymbol{E}[Z_{ij}|(\theta_j^{(t)}, b_i^{(t)}, y_{ij})] = \begin{cases} \theta_j^{(t)} - b_i^{(t)} + \dfrac{\phi(\theta_j^{(t)} - b_i^{(t)})}{1 - \Phi(-\theta_j^{(t)} + b_i^{(t)})}, & \text{se } Y_{ij} = 1, \\ \theta_j^{(t)} - b_i^{(t)} - \dfrac{\phi(\theta_j^{(t)} - b_i^{(t)})}{\Phi(-\theta_j^{(t)} + b_i^{(t)})}, & \text{se } Y_{ij} = 0. \end{cases}$$

CADEM : Passo E

Passo E2 Dadas estimativas atualizadas $\boldsymbol{b}^{(t)}$ e esperanças atualizadas $\boldsymbol{Z}^{(t+1)}$, obtidas no Passo E1, calcular

$$\widehat{\theta}_j = \boldsymbol{E}[\theta_j|(z_{ij}, b_i^{(t)}, y_{ij})] \quad = \quad \frac{\sum_{i=1}^{I}\left(z_{ij}^{(t+1)} + b_i^{(t)}\right)}{I+1}, \qquad (2)$$

Proceda da seguinte forma

Passo M Dado esperanças atualizadas $(\widehat{\boldsymbol{Z}}_{..}^{(t+1)}, \widehat{\boldsymbol{\theta}}^{(t+1)})^t$, $k = 1, 2$, atualizar $\boldsymbol{\vartheta}$ maximizando-se (1), isto é, calcular

$$\widehat{b}_i^{(t+1)} = \frac{\sum_{j=1}^n \widehat{Z}_{ij}^{(t+1)} + \sum_{j=1}^n \widehat{\theta}_j^{(t+1)} - \mu_b/\psi_b}{n + 1/\psi_b}$$

- O passo E é conduzido usando-se apenas um algoritmo numérico para calcular a f.d.a da normal padrão.
- O passo M é conduzido sem a utilização de algoritmos de maximização numérica.

Simulação

- n = 500 respondentes respondem a um teste de I = 30 itens.
- Os traços latentes foram simulados a partir de uma distribuição N (0,1).
- Os parâmetros de dificuldade do item foram escolhidos variando de (-3,3).
- Foram geradas R = 50 réplicas, ou seja, respostas dos examinados aos itens.

Comentários

- Devemos iterar os Passos E1, E2 e M até que algum critério de convergência adequado, seja alcançado.
- Estimativas dos traços latentes são obtidas, ao final do processo iterativo, através da expressão (2).
- Para calcular os erros-padrão veja o artigo original (Azevedo and Andrade (2013)).

Simulação

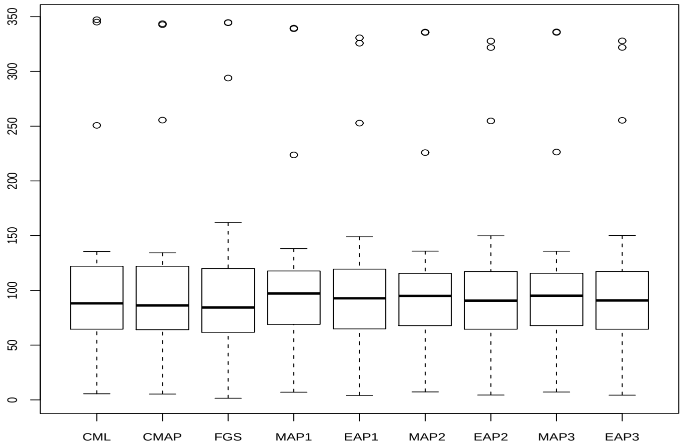
- Métodos de estimação considerados:
  - CADEM (frequentista e bayesiano) (item/traço latente).
  - Amostrador de Gibbs ("Full Gibbs sampler") (item/traço latente).
  - MVM (item).
  - MMAP (item).
  - EAP (traço latente).
  - MAP (traço latente).

Method	Item Parameter	Latent trait
CADEM ML	Maximum Likelihood CADEM	
CADEM MAP	Maximum a posterior CAEM	
FGS	Full Gibbs sampling	
MML	MML	-
MMAP (FGS prior)	MMAP with CADEM/FGS prior	-
MMAP (BILOG prior)	MMAP with Bilog-MG prior	-
MMAP MAP (FGS prior)	MMAP with CADEM/FGS prior	MAP
MMAP EAP (FGS prior)	MMAP with CADEM/FGS prior	EAP
MMAP MAP (BILOG prior)	MMAP with Bilog-MG prior	MAP
MMAP EAP (BILOG prior)	MMAP with Bilog-MG prior	EAP

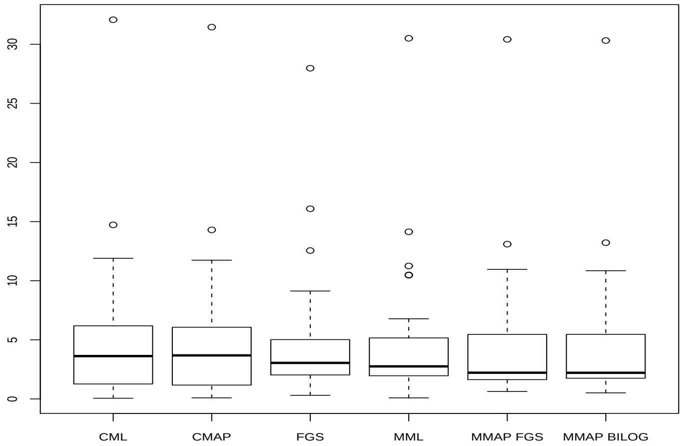
Method	Var	MeanSE	Bias	RMSE
CADEM ML	0.076	0.229	0.037	0.298
CADEM MAP	0.076	0.229	0.037	0.298
FGS	0.081	0.298	0.034	0.299
MML MAP	0.079	0.296	0.036	0.299
MML EAP	0.081	0.298	0.033	0.300
MMAP MAP (FGS prior)	0.078	0.294	0.036	0.299
MMAP EAP (FGS prior)	0.080	0.297	0.033	0.299
MMAP MAP (BILOG prior)	0.078	0.294	0.036	0.299
MMAP EAP (BILOG prior)	0.081	0.297	0.033	0.299

Method	Var	MeanSE	Bias	RMSE
CADEM ML	0.007	0.079	0.033	0.088
CADEM MAP	0.006	0.079	0.032	0.088
FGS	0.007	0.082	0.032	0.088
MML	0.007	0.081	0.030	0.088
MMAP (FGS prior)	0.007	0.080	0.029	0.087
MMAP (BILOG prior)	0.007	0.080	0.029	0.087

AVRB para as estimativas dos traços latentes



AVRB para os parâmetros de dificuldade



Comentários e conclusões

- Os passos E e M do algoritmo CADEM são conduzidos de forma direta.
- O algoritmo produziu resaultados acurados e convergência dentro de um número pequeno de iterações.
- O método de estimação proposto não apresentou problemas de convergência em nenhum caso.
- O tempo de execução computacional foi bastante pequeno.
- O modelo de 1 parâmetro se mostrou robusto com relação a ausência de normalidade.
- A estrutura do algoritmo CADEM pode ser extendida para outros MRI.

Futuras pesquisas

- Estender a abordagem proposta para outros modelos da TRI.
- Utilizar técnicas de expansão paramétrica para acelerar a convergência e melhorar a acurácia das estimativas.
- Estabelecer, analicamente (se possível), propriedades dos estimadores.
- Desenvolver uma plataforma computacional, em R, que possibilite a ampla utilização do algoritmo CADEM.

Referências

- Azevedo, C. L. N. and Andrade, D. F. (2013) CADEM: a conditional EM algorithm for fitting one parameter probit models, Brazilian Journal of Probability and Statistics, 27, 2, 245-262.
- Azevedo, C. L. N. (2003) Métodos de estimação na teroria da resposta ao item. Dissertação de Mestrado. IME-USP.
- Cai, L. (2010a). High-Dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. Psychometrika, 75, 33-57.
- Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. Journal of Educational and Behavioral Statistics, 35, 307-335.