Teoria de Resposta ao Item
Estimação

Antonio Eduardo Gomes
aegomes@unb.br

Notação

$U_{ij} = 1$ se o j-ésimo indivíduo fornece resposta positiva (ou correta) para o i-ésimo item,
e $U_{ij} = 0$ se o j-ésimo indivíduo fornece resposta negativa (ou incorreta) para o i-ésimo item,
$i = 1, \ldots, I$; $j = 1, \ldots, n$ onde
I é o número de itens no questionário, e
n é o número de respondentes (indivíduos).

Considere o ML3

$$P(U_{ij} = 1 \mid \theta_j) = c_i + \frac{1 - c_i}{1 + exp[-a_i(\theta_j - b_i)]} \, ,$$

para o qual abordaremos as possibilidades de estimação dos parâmetros dos itens (chamados de "parâmetros estruturais") e as proficiências (chamadas de "parâmetros incidentais").

Suposições

Para facilitar a estimação dos parâmetros dos itens (conhecida como "calibração"), bem como das proficiências, são assumidas as seguintes suposições:
i) unidimensionalidade do traço latente (proficiência), i.e., o teste deve ser construído de tal modo que os itens estejam medindo um único traço latente;
ii) as respostas de indivíduos diferentes são independentes;
iii) para um valor fixado do traço latente, as respostas para os diferentes itens são independentes (independência local);
Obs.: Na verdade, é possível mostrar que a unidimensionalidade implica a independência local.

Verossimilhança

Considere os vetores de parâmetros $\gamma = (\gamma_1, \ldots, \gamma_I)$ com $\gamma_i = (a_i, b_i, c_i)$, $i = 1, \ldots, I$, e $\Theta = (\theta_1, \ldots, \theta_n)$. Sejam u_{ij} valores possíveis das variáveis U_{ij}, $i = 1, \ldots, I$, e $j = 1, \ldots, n$.
A função de verossimilhança é dada por

$$\begin{aligned} L(\gamma, \Theta) &= \prod_{j=1}^{n} \prod_{i=1}^{I} [P(U_{ij} = 1 \mid \theta_j)]^{u_{ij}} [P(U_{ij} = 0 \mid \theta_j)]^{(1-u_{ij})} \\ &= \prod_{j=1}^{n} \prod_{i=1}^{I} [P(U_{ij} = 1 \mid \theta_j)]^{u_{ij}} [1 - P(U_{ij} = 1 \mid \theta_j)]^{(1-u_{ij})} \end{aligned}$$

No processo de estimação dos parâmetros, podemos ter as seguintes situações:

- As habilidades $\theta_1, \ldots, \theta_n$ são conhecidas e queremos estimar os parâmetros dos itens $\gamma_1, \ldots, \gamma_I$;
- Os parâmetros dos itens $\gamma_1, \ldots, \gamma_I$ são conhecidos e queremos estimar as habilidades $\theta_1, \ldots, \theta_n$;
- Os parâmetros dos itens e as habilidades são desconhecidos e queremos fazer a estimação conjunta de $\gamma_1, \ldots, \gamma_I$ e $\theta_1, \ldots, \theta_n$.

Em todas as situações, a estimação dos parâmetros tem que ser feita através de processos iterativos. A partir de estimativas iniciais dos parâmetros, novas estimativas são obtidas a cada iteração até obter convergência para o ponto de máximo da função de log-verossimilhança. Em casos reais, podem acontecer problemas de não convergência do algoritmo. Além disso, a convergência pode depender do ponto inicial adotado na aplicação do algoritmo.
O algoritmo mais comum para cálculo de estimativas de máxima verossimilhança (EMV) é o algoritmo de Newton-Raphson, ou o algoritmo scoring de Fisher, quando a matriz de segundas derivadas da logverossimilhança é substituída por sua esperança (i.e., pela matriz de informação com sinal invertido).
A aplicação de tais algoritmos pode, ainda, fornecer estimadores não consistentes.
Para evitar tais problemas, algoritmos alternativos foram propostos.

Parâmetros dos itens com proficiências conhecidas

Com a suposição de independência entre os indivíduos e entre os itens, podemos estimar os parâmetros de cada item separadamente. Algoritmo de Newton-Raphson: a $(k+1)$-ésima estimativa dos parâmetros dos itens é obtida a partir da k-ésima estimativa através da expressão

$$\hat{\gamma}_i^{(k+1)} = \hat{\gamma}_i^{(k)} - H^{-1}(\hat{\gamma}_i^{(k)}) \bigtriangledown \mathcal{L}(\hat{\gamma}_i^{(k)}) \,.$$

Aqui, k é o índice da iteração, H é a matriz hessiana (matriz de segundas derivadas) da log-verossimilhança com relação aos parâmetros do i-ésimo item, e $\bigtriangledown \mathcal{L}$ é o vetor gradiente (vetor de primeiras derivadas) da log-verossimilhança com relação aos parâmetros do i-ésimo item, As estimativas iniciais $\gamma_i^{(0)}$ podem ser obtidas a partir das expressões $\rho_i^{PB} = a_i / \sqrt{1 + a_i^2}$ e $\nu_i = b_i \rho_i^{PB}$, com $\Phi(-\nu_i) = \pi_i$, sendo π_i a proporção de respostas positivas para o *i*-ésimo item, ρ_i^{PB} o coeficiente de correlação ponto-bisserial para o *i*-ésimo item e $\Phi(\cdot)$ é a f.d.a. da distribuição $N(0, 1)$. A estimativa inicial de c_i pode ser dada pelo inverso do número de alternativas de respostas do item.

Esse método apresenta problema quando $u_{ij} = 0$, $j = 1, \ldots, n$. Neste caso, a verossimilhança $L(\gamma_i)$ é função extritamente crescente de b_i. Consequentemente, teríamos $\hat{b}_i = \infty$. De maneira análoga, se $u_{ij} = 1$, $i = 1, \ldots, I$, $L(\gamma_i)$ é função extritamente decrescente de b_i. Consequentemente, teríamos $\hat{b}_i = -\infty$.
Este problema pode ser evitado através da estimação bayesiana, assumindo uma distribuição *a priori* para cada parâmetro. Em geral, tomamos $a_i \sim lognormal(\mu_i, \sigma_i^2)$, $b_i \sim N(\mu_b, \sigma_b)$ e $c_i \sim Beta(\alpha_c, \beta_c)$. As estimativas bayesianas EAP (esperança a posteriori) ou MAP (moda a posteriori) podem ser obtidas via MCMC (Markov Chain Monte Carlo).

Proficiências com parâmetros dos itens conhecidos

Com a suposição de independência local e entre os itens, podemos estimar a proficiência de cada indivíduo separadamente.

Algoritmo de Newton-Raphson: a $(k+1)$-ésima estimativa da proficiência é obtida a partir da k-ésima estimativa através da expressão

$$\hat{\theta}_j^{(k+1)} = \hat{\theta}_j^{(k)} - H^{-1}(\hat{\theta}_j^{(k)})h(\hat{\theta}_j^{(k)})\,.$$

Aqui, k é o índice da iteração, H é a segunda derivada da log-verossimilhança com relação a θ_j, e h é a primeira derivada da log-verossimilhança com relação a θ_j.

A estimativa inicial $\hat{\theta}_j^{(0)}$ pode ser dada pelo escore padronizado $(T_j - m)/s$, sendo T_j o escore do j-ésimo indivíduo, m o escore médio de todos os respondentes, e s o desvio padrão do escore de todos os respondentes.

Esse método apresenta problema quando $u_{ij} = 0$, $i = 1, \ldots, I$. Neste caso, a verossimilhança $L(\theta_j)$ é função extritamente decrescente de θ_j. Consequentemente, teríamos $\hat{\theta}_j = -\infty$. De maneira análoga, se $u_{ij} = 1$, $i = 1, \ldots, I$, $L(\theta_j)$ é função extritamente crescente de θ_j. Consequentemente, teríamos $\hat{\theta}_j = \infty$.
Este problema pode ser evitado através da estimação bayesiana, assumindo uma distribuição *a priori* para a proficiência θ_j. Em geral, tomamos $\theta_j \sim N(0, 1)$. A estimativa bayesiana EAP (esperança a posteriori) ou MAP (moda a posteriori) podem ser obtidas via MCMC (Markov Chain Monte Carlo).

Neste método, assume-se uma distribuição de probabilidade $g(\theta \mid \eta)$ para a proficiência populacional, onde η representa o vetor de parâmetros desta distribuição. Comumente, adota-se a distribuição $N(0,1)$, mas outros modelos podem ser considerados. As estimativas dos parâmetros dos itens γ são dadas pelo ponto de máximo de

$$L(\gamma, \eta) = P(u_{..} \mid \gamma, \eta) = \prod_{j=1}^{n} P(u_{.j} \mid \gamma, \eta)$$

sendo

$$\begin{aligned} P(u_{.j} \mid \gamma, \eta) &= \int_{-\infty}^{\infty} P(u_{.j} \mid \theta, \gamma, \eta) g(\theta \mid \eta) d\theta \\ &= \int_{-\infty}^{\infty} P(u_{.j} \mid \theta, \gamma) g(\theta \mid \eta) d\theta \end{aligned}$$

Aqui, mesmo com a suposição de independência entre os itens, a estimação dos parâmetros dos itens tem que ser feita conjuntamente para os parâmetros de todos os itens, ou seja, o método é bastante trabalhoso computacionalmente. Utiliza-se o algoritmo de Newton-Raphson.

Mais detalhes podem ser encontrados em:

- Andrade, D.F., Tavares, H.R. e Valle, R.C., (2000). Teoria da Resposta ao Item: Conceitos e Aplicações, 14^0 SINAPE, ABE.
- Baker, F.B., Kim, S.H., (2004). Item Response Theory, Parameter Estimation Techniques, 2nd ed., Marcel Dekker.

Estimação Bayesiana

Vamos assumir as distribuições *a priori* $\gamma \sim f(\cdot \mid \tau)$ e $\theta \sim g(\cdot \mid \eta)$, e também que γ e θ são independentes, i.e., que a distribuição *a priori* conjunta de γ e θ é dada por

$$f(\theta, \gamma \mid \tau, \eta) = f(\gamma \mid \tau) g(\theta \mid \eta)$$

.

Então,

$$f(\theta, \gamma \mid u_{..}) \propto L(u_{..} \mid \theta, \gamma) f(\cdot \mid \tau) g(\cdot \mid \eta),$$

sendo $u_{..} = (u_{ij}; j = 1, \ldots, n; i = 1, \ldots, I)$, e

$$L(u_{..} \mid \theta, \gamma) = \prod_{i=1}^{I} \prod_{j=1}^{n} P(U_{ij} = u_{ij} \mid \theta, \gamma).$$

Temos que encontrar

$$
\begin{aligned}
\hat{a}_i &= \arg\max_{a_i} \{\log L(u_{..} \mid \theta, \gamma) + \log f(\gamma \mid \tau)\} \\
\hat{b}_i &= \arg\max_{b_i} \{\log L(u_{..} \mid \theta, \gamma) + \log f(\gamma \mid \tau)\} \\
\hat{c}_i &= \arg\max_{c_i} \{\log L(u_{..} \mid \theta, \gamma) + \log f(\gamma \mid \tau)\} \\
\hat{\theta}_j &= \arg\max_{\theta_j} \{\log L(u_{.j} \mid \theta_j, \gamma) + \log g(\theta_j \mid \tau)\}
\end{aligned}
$$

sendo $u_{.j} = (u_{ij}; i = 1, \ldots, I)$.

As estimativas são obtidas utilizando algoritmos de maximização (Newton-Raphson ou scoring de Fisher) para obter as soluçoes das equações:

$$
\begin{aligned}
\frac{\partial \log L(u_{..} \mid \theta, \gamma)}{\partial a_i} + \frac{\partial \log f(\gamma \mid \tau)}{\partial a_i} &= 0; \\
\frac{\partial \log L(u_{..} \mid \theta, \gamma)}{\partial b_i} + \frac{\partial \log f(\gamma \mid \tau)}{\partial b_i} &= 0; \\
\frac{\partial \log L(u_{..} \mid \theta, \gamma)}{\partial c_i} + \frac{\partial \log f(\gamma \mid \tau)}{\partial c_i} &= 0; \\
\frac{\partial \log L(u_{.j} \mid \theta_j, \gamma)}{\partial \theta_j} + \frac{\partial \log g(\theta_j \mid \tau)}{\partial \theta_j} &= 0.
\end{aligned}
$$