

# Algoritmos de seleção de itens em Testes Adaptativos Informatizados

Gilberto Pereira Sassi, Mariana Cúri  
*ICMC - USP*

## RESUMO:

Nesse trabalho, estudamos vários algoritmos de seleção de itens, ou simplesmente algoritmos/critérios de seleção, para Teste Adaptativo Informatizado (TAI). TAI's são testes seqüenciais, ou seja, os itens são aplicados um após o outro, e cada item é adaptado ao conhecimento/habilidade do respondente, medido através da Teoria de ao Item. Usamos o modelo logístico de três parâmetros da TRI e como estimador do conhecimento/habilidade usamos a média a posteriori. Fizemos um estudo de simulação, onde consideramos uma amostra de tamanho 500 da distribuição normal padrão (estamos assumindo que a nossa população segue essa distribuição). Finalmente, usamos a média e a variância do erro absoluto para concluir qual o algoritmo que apresenta melhor resultado.

**Palavras-chave:** Teste Adaptativo Informatizado, Modelo Logístico de Três parâmetros, Teoria de Resposta ao Item

## 1. INTRODUÇÃO

A pesquisa científica em áreas como Psicologia, Psiquiatria, Educação, Marketing, entre outras, tem como principal objetivo conhecer o perfil do paciente, do aluno ou do consumidor com relação a alguma característica que não é possível medir diretamente, chamada de traço latente. Para tanto, o procedimento mais utilizado é a elaboração de teste que consiste em quantificar o traço latente através do escore do teste. Contudo algumas problemas surgem. Não existe um número pré-definido de itens adequado para compor um instrumento de avaliação. Além disso, é desejável que esse número seja suficiente para avaliar níveis do traço latente baixos, médios e altos na população de interesse. Nesse contexto, alguns instrumentos podem ser longos demais, ou seja, formados por muitos itens. Este fato pode provocar cansaço no indivíduo que o está respondendo e, conseqüentemente, pode gerar respostas não confiáveis. Adicionalmente, o indivíduo é obrigado a responder itens que não estão de acordo com o seu nível de traço latente.

Nesse contexto, surgiram os testes adaptativos informatizados. Nos testes adaptativos informatizados, os itens são apresentados ao indivíduo um a um, seqüencialmente. O traço latente

do indivíduo é (re-)estimado a cada resposta dada a um item, segundo um modelo da Teoria de Resposta ao Item (TRI). Em cada passo, o item a ser apresentado é selecionado de um “banco de itens”. O escolhido deve ser aquele mais adequado para indivíduos com o nível do traço latente igual ao estimado no passo anterior. Note que diferentes indivíduos podem responder a testes distintos, com itens e/ou número de itens diferentes.

O TAI tem várias vantagens sobre os testes tradicionais. Ele gera testes menores, diminui o tempo para sua realização da avaliação, não desanima os indivíduos como ocorre quando se propõe itens muito acima ou muito abaixo de seu nível de traço latente, estima o traço latente com maior precisão, apresenta um resultado imediato. Por outro lado, os custos iniciais para seu desenvolvimento são altos. A criação de um banco de itens com os parâmetros da teoria de resposta ao item estimados é essencial, o banco deve conter vários itens apropriados para avaliar vários níveis de traço latente. Além disso, é necessário o desenvolvimento de um *software* onde o aluno responderá ao teste. Apesar das aplicações do TAI não serem isentas de problemas, suas vantagens superam as desvantagens.

## 2. METODOLOGIA

Nesta seção, apresentaremos o modelo logístico unidimensional para itens dicotômicos da TRI e alguns conceitos importantes para o desenvolvimento de um TAI.

### 2.1.1 Modelo da Teoria de Resposta ao Item

A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar acerto a um item como função dos parâmetros do item, que aqui supomos conhecidos, e da habilidade (ou traço latente) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item de Andrade *et al.* (2000). Nesse trabalho, usamos o modelo logístico de três parâmetros, dado abaixo:

$$P(U_i = 1 | \theta, \zeta_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta - b_i)}} \quad (1)$$

em que  $U_i$  é uma variável dicotômica que assume valor 1 quando  $j$  responde corretamente o item  $i$  e 0 caso contrário,  $\theta$  representa o traço latente (conhecimento) do  $j$ -ésimo indivíduo,  $\zeta_i = \{a_i, b_i, c_i\}$  é o vetor de parâmetros do item  $i$ ,  $a_i$  é o parâmetro de discriminação (ou de inclinação) do item  $i$ ,  $b_i$  é o parâmetro de dificuldade (ou de posição) do item, medido na mesma escala da dificuldade  $i$  e  $c_i$  é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responder corretamente ao item e como tal é um valor entre 0 e 1.

### 2.1.2 Estimação do traço latente

Estamos supondo que a nossa população segue a distribuição normal padrão, ou seja,  $g(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\theta^2\right)$ . A função de verossimilhança é dada por

$$L(\theta | \mathbf{u}) = \prod_{i=1}^n P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i}$$

em que  $P_i(\theta) = P(U_i = 1 | \theta, \zeta_i)$ ,  $Q_i(\theta) = 1 - P_i(\theta)$  e  $\mathbf{u} = (u_1, \dots, u_k)$ , em que  $u_j$  é o  $j$ -ésimo item selecionado pelo TAI para ser aplicado ao aluno.

Então, a distribuição a posteriori após  $k$  itens terem sido aplicados é dada por

$$g(\theta | \mathbf{u}) \propto L(\theta | \mathbf{u}) \cdot g(\theta)$$

Nesse trabalho, usamos como estimador pontual a média a *posteriori* Van Der Linden & Glas (2000).

### 2.1.3 Algoritmos de seleção

O algoritmo ou critério de seleção de item em um TAI é aquele que permite que o teste seja adaptado ao indivíduo, sendo assim o fundamento dos testes adaptativos informatizados. Abaixo apresentamos os algoritmos analisados nesse estudo.

#### Critério da máxima informação

Nesse critério, a escolha de um novo item a ser respondido pelo indivíduo, com base em suas respostas dadas a itens anteriormente apresentados, será feita através da maximização da informação de Fisher:

$$I(\theta) = \sum_{i=1}^K \frac{[\frac{d}{d\theta} P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

Neste contexto, o item selecionado será aquele que maximiza a função de Informação de Fisher. Após cada resposta, o valor de  $\hat{\theta}$  é atualizado, ou seja (re-) estimado Van Der Linden & Glas (2000).

O critério de parada do teste, isto é, o momento em que nenhum item mais é necessário ser respondido pelo indivíduo, é atingido quando o erro padrão de  $\hat{\theta}$  ( $= \frac{1}{\sqrt{I(\theta)}}$ ) estiver abaixo de um limite mínimo ou quando um número máximo de itens aplicados for atingido.

#### Critérios bayesianos de seleção

Todos os critérios bayesianos para seleção de itens envolvem alguma forma de ponderação baseada na distribuição a *posteriori* de  $\theta$ . Nessa seção faremos a suposição de que a distribuição a *priori* para  $\theta$  está bem definida. Um critério de seleção pode ser definido por

$$i_k \equiv \arg \max_j \left\{ \int_{\Theta} I_{U_j} g(\theta | \mathbf{u}) : j \in R_k \right\}$$

em que  $\mathbf{u} = \{u_1, \dots, u_k\}$ , em que  $u_j$  é o  $j$ -ésimo item selecionado pelo TAI Van Der Linden & Glas (2000).

### Critério seqüencial de Owen

Owen sugeriu escolher o  $k$ -ésimo item tal que

$$|b_k - E(\theta | \mathbf{u})| < \delta$$

para um valor pequeno de  $\delta \geq 0$ , em que  $E(\theta | \mathbf{u})$  é a esperança a *posteriori* de  $\theta$  após responder os itens  $1, \dots, k-1$  e  $b_k$  é a dificuldade do item do item  $k$  a ser selecionado no  $k$ -ésimo passo do TAI.

O critério de parada do teste é atingido quando a variância a *posteriori* alcançar um limite mínimo estipulado pelo construtor do teste ou quando um número máximo de itens respondidos pelo examinado seja atingido Van Der Linden & Glas (2000).

### Critério de seleção usando pré-*posteriori*

Os critérios bayesianos apresentados a seguir são baseados na distribuição pré-*posteriori*. Eles predizem as respostas aos itens restantes no banco depois que  $k-1$  itens terem sido aplicados e então escolhe o próximo item de acordo com as atualizações das *posteriori* para essas respostas. Um elemento chave nesta análise é a distribuição pré-*posteriori* para resposta ao item  $i$ , que tem a função de probabilidade dada por

$$P(U_{i_k} = u_{i_k} | \mathbf{U}) = \int_{\Theta} P(U_{i_k} = u_{i_k} | \theta) g(\theta | \mathbf{u}) d\theta.$$

Suponha que o item  $i$  seja selecionado. O examinado responde corretamente a este item com probabilidade  $P(1 | \mathbf{u})$  ou incorretamente com probabilidade  $P(0 | \mathbf{u})$ , em que  $\mathbf{u} = \{u_1, \dots, u_{k-1}\}$ . Qualquer resposta conduz a atualização da distribuição a *posteriori*, respectiva variância a *posteriori* e estimação pontual.

Um critério baseado na análise de predição é o critério do risco pré-posterior sobre a perda quadrática Van Der Linden & Glas (2000),

$$i_k \equiv \arg \min_j \{P(U_j = 0 | \mathbf{u}) \text{Var}(\theta | \mathbf{u}, U_j = 0) + P(U_j = 1 | \mathbf{u}) \text{Var}(\theta | \mathbf{u}, U_j = 1)\}$$

### 3 Resultados

Esse estudo foi realizado por meio de simulações em R. Para esse fim, usaremos uma amostra de tamanho 500 da distribuição normal padrão. Esse estudo será realizado para três tamanhos distintos de banco de itens:

- (i) Um banco de tamanho pequeno, com cinquenta itens. Para esse tamanho de banco, usaremos como critério de parada a precisão da estimativa ou até que um número máximo de itens aplicados seja alcançado. Por precisão da estimativa entendemos o momento em que a recíproca da raiz da Informação de Fisher (para critérios de seleção clássicos) ou a variância a *posteriori* (para critérios bayesianos) esteja abaixo de um valor pré-estabelecido.
- (ii) Um banco de tamanho médio, com 300 itens. Para esse tamanho de banco, usaremos como critério de parada a precisão da estimativa ou até que um número máximo de itens aplicados seja alcançado.
- (iii) Um banco grande, com 700 itens. Para esse tamanho de banco, usaremos como critério de parada a precisão ou até que um número máximo de 10, 20 e 30 itens aplicados seja alcançado.

Nas simulações, a resposta a cada item selecionado pelo critério em questão será obtida por meio de uma amostra de tamanho 1 da distribuição bernoulli com probabilidade de resposta correta dada pelo modelo logístico de três parâmetros em que  $\theta$  é o valor verdadeiro da habilidade (um dos 500 valores da amostra da distribuição normal padrão adotada nesse estudo). Após cada resposta, a estimativa do traço latente é atualizada.

Compararemos os resultados através de uma adaptação das seguintes medidas

$$\hat{\sigma}^2 = \frac{1}{500} \sum_{i=1}^{500} |\theta_i - \hat{\theta}_i|^2 \quad \text{e} \quad \hat{\mu} = \frac{1}{500} \sum_{i=1}^{500} |\theta_i - \hat{\theta}_i|$$

em que  $\theta_i$  é o verdadeiro valor da habilidade,  $\hat{\theta}_i$  é o valor estimado pelo TAI,  $\hat{\sigma}^2$  é variância do erro absoluto e  $\hat{\mu}$  é média do erro absoluto.

### Referências

- de Andrade, D., Tavares, H. & da Cunha Valle, R. (2000). Teoria da resposta ao item: conceitos e aplicacoes. *ABE, Sao Paulo*.
- Van Der Linden, W. & Glas, C. (2000). *Computerized adaptive testing: Theory and practice*. Springer Netherlands.