versão impressa ISSN 0101-7438 / versão online ISSN 1678-5142

# ANÁLISE BAYESIANA DO FUNCIONAMENTO DIFERENCIAL DO ITEM

**Tufi Machado Soares***
Departamento de Estatística
Centro de Avaliação Educacional (CAEd)
Universidade Federal de Juiz de Fora (UFJF)
tufi@caed.ufjf.br

**Dani Gamerman**
Departamento de Métodos Estatísticos
Universidade Federal do Rio de Janeiro (UFRJ)
dani@im.ufrj.br

**Flávio Bambirra Gonçalves**
Departamento de Métodos Estatísticos
Universidade Federal do Rio de Janeiro (UFRJ)
flavio@dme.ufrj.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

*Recebido em 06/2006; aceito em 04/2007*
*Received June 2006; accepted April 2007*

**Resumo**

Neste trabalho utiliza-se a abordagem Bayesiana na estimação dos parâmetros de modelos da Teoria da Resposta ao Item, destinados à análise do Funcionamento Diferencial do Item, DIF – *Differential Item Functioning*. Os modelos propostos são integrados, e permitem incorporar estruturas de regressão que podem ser usadas para explicar o DIF relacionado à co-variáveis associadas aos itens. São considerados modelos para múltiplos grupos e a abordagem utilizada incorpora naturalmente o erro de estimação do traço latente e dos parâmetros estruturais. A abordagem permite, naturalmente, considerar DIF tanto na dificuldade quanto na discriminação do item. Exemplos com dados simulados e com dados reais são apresentados.

**Palavras-chave:** funcionamento diferencial do item; MCMC; teoria da resposta ao item.

**Abstract**

This paper uses a Bayesian approach for parameter estimation in Item Response Theory Models for DIF – Differential Item Functioning – analysis. The models proposed are integrated, and incorporate regression structures that can be used to explain the DIF related to items associated covariates. The models are proposed for multiple groups and the approach used, naturally, consider the estimation error of the latent trace and the estimation error of the structural parameters. Examples with simulated data and real data are also presented.

**Keywords:** item differential functioning; MCMC; item response theory.

## 1. Introdução

Informalmente, pode-se entender que um item de um teste de avaliação educacional apresenta um funcionamento diferencial (DIF – *Differential Item Functioning*) para dois grupos de alunos quando apesar de esses alunos terem o mesmo nível de proficiência ou de habilidade cognitiva e, ainda assim, o desempenho no item é significativamente diferente nos dois grupos. Por exemplo, um tipo de funcionamento diferencial muito comum é aquele em que o item é mais fácil para um determinado grupo de alunos do que para outro, mesmo quando são comparados alunos de mesma proficiência.

Num primeiro momento, não é desejável encontrar itens que apresentam DIF num instrumento de avaliação educacional, pois isso poderia implicar que algum particular grupo de indivíduos esteja sendo privilegiado em detrimento dos demais. Esse é o contexto histórico da preocupação com o DIF, que estava ligado ao desejo de se construir questões de testes que não fossem afetadas por características étnico-culturais dos grupos submetidos a testes de admissão (*cf.* Cole, 1993). Muito associado, portanto, às campanhas em prol da melhoria dos direitos civis dos cidadãos comuns nos anos 60 nos Estados Unidos da América. Inicialmente, então, o foco dos estudos era o de identificar tipos de itens que eram favoráveis (ou desfavoráveis) a um determinado grupo em detrimento de outros, de tal forma que os testes pudessem evitar questões "prejudiciais e injustas".

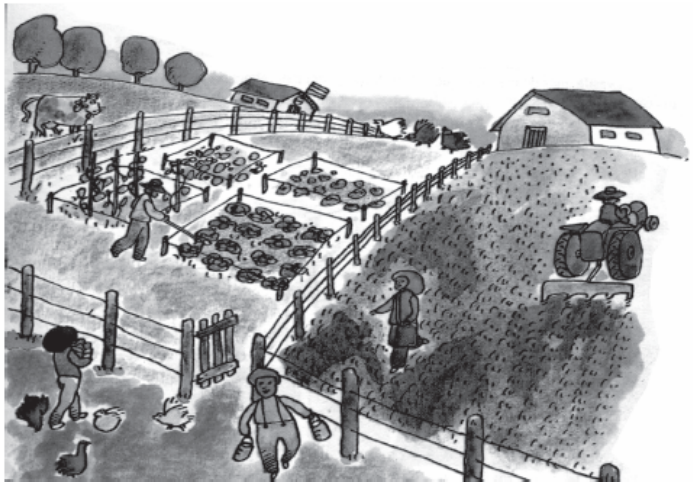
No entanto, os itens que apresentam DIF trazem informações importantes e nem sempre o procedimento adotado atualmente é, simplesmente, de retirá-los das provas ou dos cálculos das proficiências – desde que o comportamento diferencial não tenha impacto apreciável sobre a proficiência estimada. Por outro lado, tendo em vista que é impossível eliminar completamente a presença de itens com algum tipo de funcionamento diferencial, pois é impossível considerar todas as diferenças entre os alunos, é razoável supor que sempre haverá itens com algum tipo de funcionamento diferencial em um teste educacional. Note-se que, embora o natural seja admitir que grupos diferentes sejam constituídos a partir da localização dos alunos no espaço e no tempo, ou por diferenças de características étnico-culturais e ou biológicas, os grupos podem ainda ser constituídos a partir de outras características como, por exemplo, por diferenças nos métodos de ensino aos quais os alunos são expostos. Assim, itens com DIF podem vir a ser introduzidos propositalmente para testar hipóteses sobre diferenças cognitivas entre diferentes grupos de alunos, por exemplo.

A Figura 1 corresponde à imagem de um item (cedido pela Secretaria Estadual de Educação de Minas Gerais) que foi aplicado aos alunos da 4ª série do ensino fundamental no PROEB/ SIMAVE-2001, programa de avaliação do ensino fundamental e médio da rede pública estadual de Minas Gerais, que avaliou competências nas disciplinas de história e geografia.

Através do Sistema de Análise de Itens (SisAni), desenvolvido em linguagem DELPHI (*cf.* Soares & Galvão, 2004), foram calculadas as estatísticas clássicas do funcionamento diferencial e foi produzida a Figura 2 que mostra o funcionamento do item para dois grupos de alunos. As estatísticas e os gráficos apresentados permitem uma análise do comportamento diferencial com respeito, principalmente, à dificuldade do item. Não sendo relevante neste ponto, sugere-se ao leitor não familiarizado ignorar as estatísticas e concentrar sua atenção na interpretação dos gráficos. No primeiro deles é apresentado o percentual de acerto do item para os alunos de ambos os grupos. A comparação é apresentada dividindo-se os alunos em subgrupos pareados de mesma habilidade cognitiva, isto é, cada quadrado nesse gráfico representa o percentual de acerto para os alunos de um dos grupos (denominado grupo de referência) com proficiências em torno de um determinado nível e o triângulo

correspondente representa o percentual de acerto dos alunos do outro grupo (denominado grupo focal) com proficiência em torno desse mesmo nível. O segundo gráfico apresenta a diferença entre esses percentuais de acertos calculados para os dois grupos.

(PROEB-2001) Observe o desenho e responda às duas questões seguintes:



(H04038MG) O desenho mostra uma realidade que se relaciona a um espaço:

A) urbano.
B) industrial.
C) rural.
D) metropolitano.

**Figura 1** – Imagem do Item H04038MG (SIMAVE/PROEB – 2001).

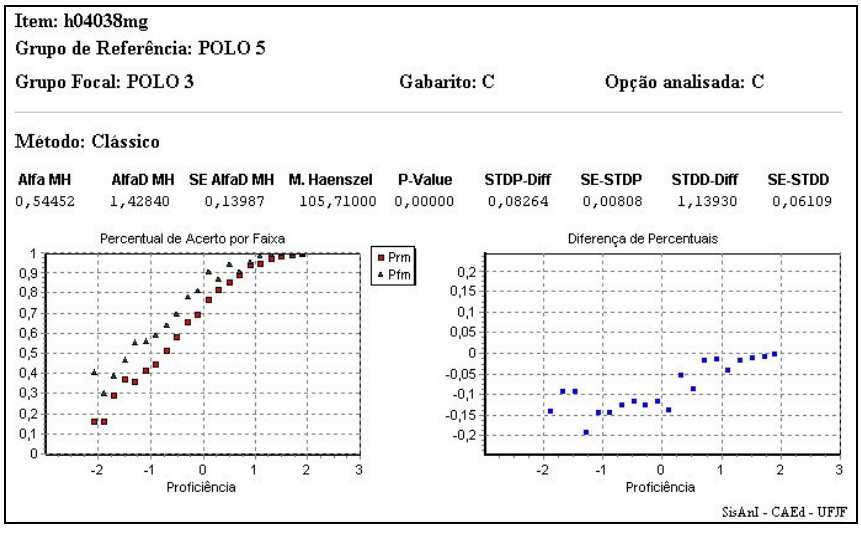



**Figura 2** – Funcionamento Diferencial exibido pelo Item H04038MG.

No caso são comparados os desempenhos dos alunos da região metropolitana de Belo-Horizonte (Grupo de Referência) com os desempenhos dos alunos da região do triângulo mineiro (Grupo Focal). Observa-se que os alunos da região do triângulo acertaram o item em maior proporção do que os alunos da região metropolitana, ou seja, o item foi mais fácil para os alunos do triângulo mineiro do que para os alunos da região metropolitana. O valor para a estatística *AlfaD MH* (alfa/Delta de Mantel Haenszel – ver Dorans & Holland, 1993, para a definição formal dessa estatística) de 1,428 indica um nível intermediário de funcionamento diferencial (pelo critério da *ETS – Educational Testing Service* – ver Longford, Holland & Thayer, 1993, por exemplo).

O objetivo deste trabalho é o de propor e implementar um novo modelo para auxiliar uma análise de DIF. Esse novo modelo fornece elementos para uma decisão baseada em estatísticas quanto às conclusões sobre a natureza do DIF. Isto é alcançado a partir da introdução de co-variáveis explicativas no modelo que representam determinadas características dos itens. Embora algumas abordagens já tenham sido propostas e implementadas nesse sentido (ver Swanson *et al.*, 2002), até agora, nenhuma delas propôs e implementou um modelo com a estimação simultânea das proficiências e dos parâmetros estruturais, inclusive com os parâmetros associados às co-variáveis explicativas. Nesse sentido, o modelo proposto e implementado neste trabalho é flexível e permite uma abordagem integrada. Pelo fato de o modelo apresentar uma estrutura complexa, utiliza-se para a estimação dos parâmetros técnicas de simulação de Monte Carlo conhecidas como *MCMC* (*Markov Chain Monte Carlo*).

Na seção 2 será apresentado o modelo proposto para análise de DIF. Na seção 3, apresenta-se uma rápida revisão da literatura sobre o assunto, contextualizando o presente trabalho. Na seção 4 será apresentado o método utilizado para estimação do modelo, e na seção 5 serão apresentados exemplos, a partir de dados simulados e dados reais. Finalmente, na seção 5 são apresentadas conclusões finais do trabalho.

## 2. Modelo para a identificação e análise do DIF

Modelos da teoria da resposta ao item associam a probabilidade de o aluno alcançar um determinado escore no item com sua habilidade latente, ou proficiência, $\theta_j$ (ver, por exemplo, Lord (1980) para uma discussão sobre esse conceito). O modelo de três parâmetros, proposto por Birnbaum (1968), tem tido um importante papel no contexto da avaliação educacional em larga escala, principalmente, porque nesses casos geralmente se empregam itens de múltiplas escolhas nos testes e o efeito de um acerto devido a uma escolha, pelo menos em parte, aleatória é considerado com a introdução do parâmetro $c_i$ no modelo de dois parâmetros originalmente proposto (ver Lord, 1980). O gráfico que representa a relação entre a probabilidade de acerto e a proficiência é conhecido como Curva Característica do Item (CCI). Pode-se ter, então, a CCI obtida a partir do modelo (CCI teórica) ou, se são conhecidas as proficiências dos alunos, a CCI construída a partir dos dados empíricos (CCI empírica).

Tipicamente, em avaliação educacional, um teste é constituído por $I$ itens, mas um aluno $j$ responde apenas a um subconjunto $I(j)$ desses itens. Seja então $Y_{ij}$, $j = 1,...,J$, o escore atribuído à resposta dada pelo aluno $j$ ao item $i$, $i \in \mathrm{I}(j)$, $(\mathrm{I}(j) \subset [1,...,I])$. Neste trabalho vai-se considerar apenas o caso dicotômico, onde ao item é atribuído um dos escores [0,1], de tal forma que $Y_{ij} = 1$ representa um acerto, por exemplo, e $Y_{ij} = 0$ representa um erro.

Em geral, pode-se ter diferentes tipos de DIF (ver, por exemplo, Hanson, 1998, para uma caracterização mais abrangente), mas restringindo-se às características explicitadas através do modelo de três parâmetros, pode-se imediatamente caracterizar o tipo de DIF de acordo com a dificuldade, a discriminação e o acerto casual. Dessa forma, a análise do DIF consiste em verificar a estabilidade do modelo nos diferentes grupos de alunos, isto é, verificar se os parâmetros dos modelos dos itens são diferentes ou não para os grupos.

Neste trabalho não se vai incluir a possibilidade de DIF no parâmetro de acerto casual do item. Embora seja possível, as dificuldades conhecidas para a estimação desse parâmetro e restrições de ordem práticas limitam substancialmente a aplicabilidade de uma implementação nesse sentido.

O modelo proposto neste trabalho para a análise do DIF é apresentado através da equação (1) e da equação (2).

$$P(Y_{ij}=1|\theta_j,a_i,b_i,c_i,d^a_{gi},d^b_{gi}) = c_i + \frac{(1-c_i)}{1+\Delta}, \qquad (1)$$

$$\Delta = e^{-D\, e^{d^a_{gi}}\, a_i\,(\theta_j - b_i + d^b_{gi})}, \qquad g=1,...,G.$$

A equação (1) representa a probabilidade de um aluno j acertar ao item i do teste dado que ele tenha a proficiência $\theta_j$. Os demais parâmetros representam as características do item. O parâmetro $a_{gi} = e^{d^a_{gi}}\, a_i\ (>0)$ é conhecido como o parâmetro de discriminação do item, $b_{gi} = b_i - d^b_{gi}$, como o parâmetro de dificuldade do item e $c_i\ (\in [0,1])$, como o parâmetro de acerto casual (Lord, 1980, por exemplo, apresenta justificativas para essas denominações). Neste modelo admite-se, ainda, que os alunos estejam agrupados em $G$ grupos e que $\theta_j \sim N(\mu_{g(j)},\sigma_{g(j)})$, onde $g(j)$ identifica grupo do aluno j. Para garantir a identificabilidade do modelo admite-se que $\lambda_1 = [\mu_1\ \sigma_1] = [0\ 1]$ – considerado como grupo de referência. Por outro lado, os demais parâmetros das distribuições de proficiências, representados por $\lambda_g := [\mu_g\ \sigma_g]$, $g=2,...,G$, é desconhecido e deverá ser estimado em conjunto com os demais parâmetros. O parâmetro $d^b_{gi}$ ($d^b_{1i}=0$) representa o DIF com relação à dificuldade do item para cada grupo e o parâmetro $d^a_{gi}$ ($g=2,...,G$, $d^a_{1i}=0$) representa o DIF com relação à discriminação.

Para garantir a comparabilidade das proficiências estimadas para os alunos dos diferentes grupos, é necessário que parte dos itens que sejam aplicados em comum aos grupos não deva apresentar DIF (no caso do modelo acima isto implica que $d^a_{gi} = d^b_{gi} = 0$ para esses itens). Thissen, Steinberg & Wainer (1993), denomina esses itens que não exibem DIF e são usados para garantir a comparabilidade das proficiências dos alunos dos diferentes grupos de itens âncoras (*Anchor Itens*). Note-se que o subconjunto de itens âncoras deve ser definido a priori. Neste trabalho, ele será representado por $I_A \subset [1...I]$.

O conjunto dos itens comuns para os quais se deseja analisar o DIF será representado por $I_{dif} = \{1,...,I\} - I_A$. Naturalmente, para alguns itens, pode-se admitir o DIF com relação à dificuldade e não com relação à discriminação ou vice-versa. Aliás, o mais comum é a presença de DIF com relação apenas à dificuldade. Nesse sentido, o conjunto $I^a_{dif} \subset I_{dif}$, representa o conjunto de itens para os quais se admite DIF na discriminação e, o conjunto $I^b_{dif} \subset I_{dif}$, representa o conjunto correspondente para o DIF na dificuldade.

Finalmente, para investigar a natureza do DIF impõe-se uma estrutura de regressão para $d^h_{gi}\,(h = a,b)$:

$$d^h_{gi} = \gamma^h_{0g} + \sum_{k=1}^{K} \gamma^h_{kg}\ W^h_{ki} + \eta^h_{gi} \qquad (2)$$

onde $\gamma^h_{kg}$ são os parâmetros fixos do modelo de DIF, $W^h_{ki}$ são variáveis explicativas associadas aos itens e $\eta^h_{gi}$ representa um fator aleatório específico do item em cada grupo. Vai-se admitir que $E\left(\boldsymbol{\eta}^h_g\right) = 0$ e $E\left(\eta^h_g \left(\eta^h_g\right)^T\right) = T$, ou $E\left(\eta^h_g \left(\eta^h_g\right)^T\right) = I\ \tau^2_g,\ \forall g = 2,...,G$, sendo *I* a matriz identidade.

A função de verossimilhança é dada, então, por:

$$f(Y \mid \theta, a, b, c, d^a, d^b) = \prod_{j=1}^{J} \prod_{i=1}^{I(j)} \left( P(Y_{ij} = 1 \mid \theta_j, a_i, b_i, c_i, d^a_{gi}, d^b_{gi}) \right)^{Y_{ij}} \left(1 - P(Y_{ij} = 1 \mid \theta_j, a_i, b_i, c_i, d^a_{gi}, d^b_{gi})\right)^{1-Y_{ij}},$$

Como o objetivo do trabalho é apresentar uma análise Bayesiana para o problema do DIF, o modelo se completa com a especificação das distribuições a priori para os parâmetros. As prioris adotadas para os parâmetros estruturais foram: $a_i \sim LN(0,2)$, $b_i \sim N(0,1)$ e $c_i \sim beta(5,17)$. Estas prioris são habitualmente empregadas, como por exemplo, são *defaults* no *software Bilog-mg*, e naturais tendo em vista as características dos parâmetros. Para os parâmetros correspondentes ao modelo de explicação do DIF, admite-se que $\mathbf{d}^h_g \mid \mathbf{W}^h, \boldsymbol{\gamma}^h_g, \mathbf{T}^h \sim N(\mathbf{W}^h \boldsymbol{\gamma}^h_g, \mathbf{T}^h)$, com a priori $\gamma^h_g \sim N(\gamma^h_0, S^h_0)$. Admite-se, ainda, que $\mu_g \mid \sigma_g \sim N(0, \sigma_g)$, com $\sigma^2_g \sim GI(\alpha_g, \beta_g)$, onde *GI* representa a distribuição Gama Inversa.

## 3. Abordagens Tradicionais para a Detecção e Análise do DIF

Lord (1980) propôs utilizar um teste normal para verificar se há diferenças significativas entre os parâmetros de dificuldade, e um teste qui-quadrado para verificar se há diferenças nos parâmetros de dificuldade e discriminação simultaneamente. Lord (1980) não considerou DIF no parâmetro de acerto casual. Nessa abordagem, as estatísticas serão eficientes na detecção do DIF se, ou a proficiência já for conhecida para os alunos, ou se houver um bom número de itens âncoras de tal forma que as proficiências sejam estimadas com erros muito

pequenos. Ora, admitir que a proficiência seja conhecida é sempre uma ficção e, em geral, o número de itens respondido por um aluno é pequeno. Assim, propriedades assintóticas podem ser pouco confiáveis dependendo do tipo de estimador utilizado – Lord utilizou originalmente um estimador de máxima verossimilhança conjunta para os parâmetros estruturais do modelo e para as proficiências dos alunos, o que é inconsistente (*cf.* Baker, 1992). Estudos mostraram que o método de Lord pode conduzir a erros bastante expressivos na detecção do DIF (ver Thissem, Steinberg & Wainer, 1993).

Thissem, Steinberg & Wainer (1993) apresentaram um procedimento geral para detecção de DIF que denominaram de método geral TRI-Razão de Verossimilhanças ("general IRT-LR"). *Grosso modo*, o método propõe que o conjunto de itens âncoras seja cuidadosamente selecionado por testes anteriores e especialistas – *Designated Anchor*. Os autores propõem, inicialmente, maximizar a função de verossimilhança marginalizada com respeito à distribuição de proficiências considerando-se que todos os itens não apresentam DIF e os alunos pertencem ao mesmo grupo. Numa segunda etapa, maximiza-se a função de verossimilhança marginalizada com respeito às distribuições de proficiências considerando-se que parte dos itens pode apresentar DIF, e os alunos pertençam a grupos múltiplos. Os dois modelos, então, são comparados através de um teste de razão de verossimilhanças. O *software* Bilog-mg tem implementado um algoritmo nessa linha de abordagem, estimando diferentes valores para os parâmetros de dificuldade nos diferentes grupos $\left(b_{gi}\right)$, mas não permite, pelo menos diretamente, definir um subconjunto de itens âncoras, de tal forma que a restrição $\sum_{i=1}^{I}\left(b_{0i}-b_{gi}\right)=0$ faz-se necessária para garantir a identificabilidade. Essa restrição impede, ou pelo menos restringe, uma análise de DIF mais abrangente. Por outro lado, a consistência do estimador de máxima verossimilhança marginal depende do conhecimento das distribuições corretas. Apesar de o programa estimar em etapas iterativas os valores da distribuição nos pontos de quadratura pelo procedimento proposto em Bock & Zimowski (1997), são desconhecidos estudos que garantam uma boa convergência da distribuição de proficiências. Esta, provavelmente, depende do número de itens sem DIF e da quantidade de informação associada a eles sobre as proficiências. Por outro lado, o procedimento proposto por Thissem, Steinberg & Wainer (1993) tem uma preocupação muito mais focada no impacto que o DIF pode ter sobre a proficiência do que propriamente numa análise de DIF. Por exemplo, estes autores não introduziram co-variáveis explicativas nos modelos.

O método da regressão logística para identificar o DIF foi proposto por Swaminathan & Rogers (1990), e, basicamente**,** admitindo que a proficiência é conhecida, utiliza o modelo de regressão logística:

$$P(Y=1|\theta,\beta;g)=\frac{1}{1+e^{-D\,a_i(\theta-b_i+d_i^b\,g+d_i^a\,g\,\theta)}},\quad g=0,1.$$

Nesse caso, os grupos são representados pela variável $g$ ($g=0$ se o examinando pertence ao grupo de referência e $g=1$ se o examinando pertence ao grupo focal). Evidentemente, o modelo pode ser generalizado, introduzindo-se outras variáveis *dummy*, para comparação com mais grupos.

Os parâmetros dos modelos são estimados pelos métodos habituais e testes de significância (tipicamente baseados na estatística de Wald) indicarão a existência de DIF quanto à

dificuldade e quanto à discriminação, respectivamente. Alguns autores utilizam o escore bruto do teste, ou variações, como, por exemplo, o escore bruto excluindo-se o item cujo DIF está sendo testado.

Note-se que o modelo da regressão logística é aparentemente similar ao proposto através da equação (1), no entanto, ele não considera o acerto casual ao item e, principalmente, admite que a proficiência seja conhecida. É um artifício que simplifica a estimação dos parâmetros estruturais do modelo, porém, a qualidade dos resultados depende da qualidade da medida de proficiência disponível.

Swanson *et al.* (2002) propõem uma extensão do modelo de regressão logística que emprega estruturas hierárquicas segundo os itens. As características específicas dos itens são representadas, então, por co-variáveis explicativas, que podem ser variáveis indicadoras, como por exemplo se o item está associado a uma particular competência, ou variáveis intervalares, como o número de palavras empregadas no enunciado. Mas a restrição de proficiência conhecidas continua sendo necessária naquele trabalho.

Patz & Junker (1999b) discutem e apresentam aplicações de MCMC na teoria da resposta ao item. De passagem, os autores comentam a viabilidade de se implementar através dessa técnica um modelo para DIF que incluiria co-variáveis associadas aos itens. No entanto, eles não implementam qualquer modelo nesse sentido naquele trabalho nem em qualquer outro a posteriori que seja do conhecimento destes autores. A idéia neste artigo é estimar, nas mesmas linhas propostas em Patz & Junker (1999a) e Patz & Junker (1999b), os parâmetros do modelo apresentado em (1) e (2) que considera a possibilidade de DIF incluindo co-variáveis explicativas associadas.

## 4. Estimação do Modelo Através de MCMC

O número de parâmetros do modelo formulado em (1) e (2), e as diferentes características desses parâmetros, torna sua estimação consideravelmente difícil, em particular, quando a proficiência não é conhecida. A distribuição a posteriori conjunta dos parâmetros não apresenta forma fechada e é de difícil tratamento por métodos numéricos. O método de estimação dos parâmetros dos modelos que foi adotado neste trabalho consiste em encontrar a média da distribuição conjunta a posteriori de todos os parâmetros dos modelos dos itens e proficiências dos alunos condicionados aos dados, isto é encontrar o valor esperado da distribuição:

$$p(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{T} \mid \mathbf{Y}, \mathbf{W}), \ \boldsymbol{\beta} = [\mathbf{a}\ \mathbf{b}\ \mathbf{c}], \ \mathbf{d} = [\mathbf{d}^{\mathbf{a}}, \mathbf{d}^{\mathbf{b}}].$$

Para superar as dificuldades mencionadas, empregam-se técnicas de simulação extensivas conhecidas como *MCMC – Markov Chain Monte Carlo*, que permitem gerar amostras dessa distribuição conjunta.

Em particular, foi utilizado o amostrador de Gibbs (*cf.* Gamerman & Lopes, 2006) que consiste num algoritmo iterativo que permite gerar amostras de uma distribuição conjunta a posteriori, como a apresentada acima, a partir de amostras obtidas das distribuições de cada parâmetro ou de bloco de parâmetros condicionadas aos demais parâmetros dos modelos, chamadas de distribuições condicionais completas. Sob condições apropriadas, mostra-se que a seqüência de variáveis aleatórias que representam as sucessivas amostras geradas constituem uma cadeia de Markov cuja distribuição de transição converge para uma

distribuição de transição estacionária e igual à distribuição conjunta a posteriori dos parâmetros.

Quando as distribuições condicionais completas não apresentam formas fechadas conhecidas torna-se necessário utilizar, por exemplo, o algoritmo de Metropolis-Hastings (*cf.* Gamerman & Lopes, 2006). Nesse caso, a cada iteração do algoritmo uma amostra de um parâmetro é gerada a partir de uma distribuição arbitrária com parâmetro de localização (tipicamente, a média) dado pela amostra anterior do parâmetro. Naturalmente, essa distribuição deve ser adequada às restrições impostas ao parâmetro e é denominada de transição proposta. A nova amostra é aceita ou rejeitada (mantendo-se, neste caso, o estado anterior) a partir de uma decisão tomada de acordo com uma determinada probabilidade (denominada de probabilidade de aceitação).

As distribuições condicionais completas utilizadas na geração das amostras dos parâmetros dos modelos dos itens são apresentadas em detalhes no anexo A.

## 5. Desempenho em Exemplos Simulados e Estudos Reais

Exemplo 1 (2 grupos – DIF na dificuldade)

Neste primeiro exemplo, foram simuladas as respostas de 4000 alunos a um teste constituído de 50 itens dicotômicos – os alunos foram divididos em dois grupos com 2000 alunos cada. Os parâmetros dos modelos dos itens e as proficiências dos alunos foram todos gerados aleatoriamente. As proficiências foram geradas a partir de uma distribuição normal, com média zero e desvio-padrão 1,0 para o grupo de referência e média 0,15 e desvio-padrão 1,0 para o grupo focal. Os parâmetros dos modelos foram gerados de acordo com as prioris apresentadas na seção 2. O programa de simulação escolhe aleatoriamente os itens âncoras, neste exemplo 33 dos itens foram escolhidos como âncoras e, portanto, os outros 17 itens foram considerados, então, como sujeitos a um possível funcionamento diferencial. Foram fixados os valores $\gamma_{02} = 0,3$ e, $\gamma_{12} = 0,3$, e $W_{1i} = 1$ para 7 desses itens e $W_{1i} = 0$ para os demais, simulando assim o efeito de uma co-variável binária associada a uma possível característica comum desses sete itens. Os parâmetros $d_{2i}^b$ foram gerados conforme o modelo em (2) supondo que $\eta_{1i}^b \sim N(0,0,2)$ e, $\tau_g^2 = 0,04$.

Todos os estudos foram realizados com o auxílio de dois programas elaborados pelos autores em duas linguagens diferentes: R e Matlab. Os resultados são os mesmos obtidos em ambos os programas, porém o programa em Matlab gasta cerca de um terço do tempo gasto pelo programa em R. Nesse exemplo, com 2 grupos, o programa em Matlab, executado em um computador PENTIUM IV – 2,6 GHz, gastou cerca de 8 horas para a geração das 20000 realizações das cadeias.

As convergências das cadeias de Markov geradas pelo amostrador de Gibbs foram testadas através do critério R de Gelmam & Rubin (RGR) (*cf.* Gameman & Lopes, 2006), a partir da geração de 4 cadeias em paralelo, de 20000 realizações cada, com diferentes condições iniciais. Todas as cadeias alcançaram convergência com menos de 10000 realizações, apresentando RGR < 1.1. A Figura 3 mostra as 10000 primeiras realizações encontradas para os parâmetros estruturais do item 1, além dos parâmetros correspondentes as médias dos grupos e os parâmetros da estrutura de regressão sobre o DIF (ver Figura 4).

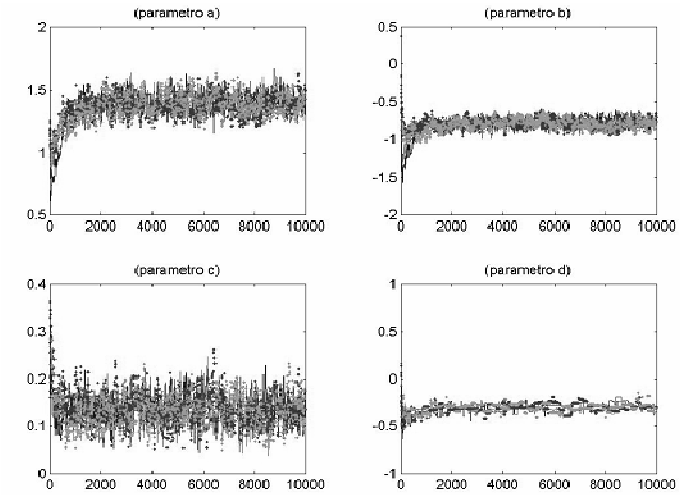



**Figura 3** – Trajetória das 4 Cadeias para os Parâmetros Estruturais Selecionados.

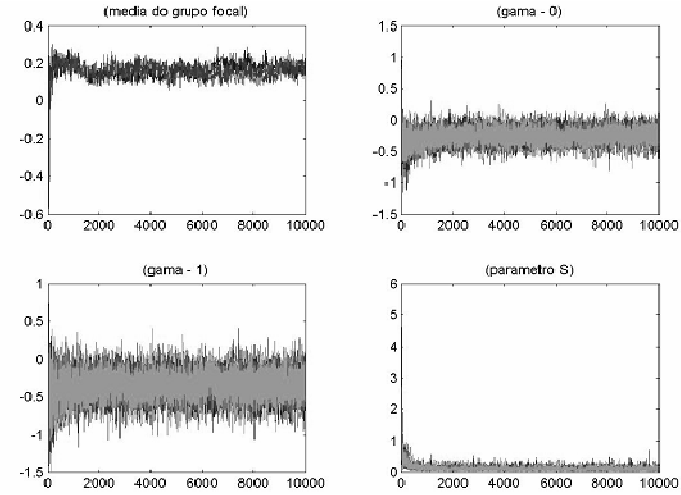



**Figura 4** – Trajetória das 4 Cadeias para os Parâmetros de Regressão.

Na Tabela I são apresentadas as estimativas que foram obtidas para a média e o desvio padrão das proficiências do grupo 2. Foram utilizadas as médias das 10000 últimas realizações das cadeias.

**Tabela I** – Média e desvio-padrão dos Grupos.

| *Grupo* | $\mu_g$ *(Real)* | $\bar{\mu}_g$ *(Est.)* | *Intervalo de Credibilidade (I.C.) (95%)* | | $\sigma_g$ *(Real)* | $\bar{\sigma}_g$ *(Est.)* | *Intervalo de Credibilidade (I.C.) (95%)* | |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 2 | 0,1500 | 0,1616 | 0,0898 | 0,2296 | 1,0000 | 1,0341 | 0,9701 | 1,1819 |

Verifica-se que tanto a estimativa da média quanto do desvio padrão são muito satisfatórias, bastante próximas dos valores reais. Esse resultado garante, portanto, que as médias de proficiências dos dois grupos de alunos estão sendo apropriadamente comparadas mesmo quando há itens com funcionamento diferencial.

Em geral, como pode ser visto na Tabela II, a estimação dos parâmetros estruturais a e b foi muito satisfatória, enquanto a estimação do parâmetro c é, em geral, menos satisfatória. A correlação de Pearson entre os parâmetros verdadeiros e as estimativas foi de, respectivamente, 0,9750, 0,9970 e 0,6020.

Problemas foram encontrados na estimação dos parâmetros de itens que exibiam ou valores muito baixos ou muito elevados para os parâmetros de discriminação. No entanto, nessas condições, é naturalmente mais difícil se obter boas estimativas.

Nota-se que os intervalos de credibilidade são muito menores, proporcionalmente aos valores, para os parâmetros de dificuldade e muito maiores para os parâmetros de acerto casual. Esse fato indica que os estimadores dos parâmetros de dificuldade são muito mais precisos do que os estimadores dos parâmetros de acerto casual. Em alguns exemplos simulados, sem DIF, as estimativas obtidas através do amostrador de Gibbs – para todos os três parâmetros do modelo – têm sido superiores às obtidas pelo *software Bilog-mg* que utiliza métodos de máxima verossimilhança marginal e máxima verossimilhança marginal penalizada (MMAP) – ou máxima distribuição a posteriori marginalizada. A estimação dos parâmetros de DIF $d_{2i}^{b}$ também se mostrou bastante satisfatória, a correlação de Pearson entre os parâmetros verdadeiros e as estimativas foi de 0,9942. Os resultados apresentados na Tabela III indicam que a utilização do intervalo de credibilidade pode ser útil para verificar se o item apresenta ou não um DIF significativo. Por exemplo, os itens 14 e 42 apresentam, claramente, um DIF desprezível.

**Tabela II** – Parâmetros DIF.

| *Item* | $d_{2i}^{b}$ *(Real)* | $\bar{d}_{2i}^{b}$ *(Est.)* | *I. C. (95%)* | |
|---|---|---|---|---|
| 1 | -0,3066 | -0,3033 | -0,3908 | -0,2262 |
| 2 | -0,2695 | -0,2936 | -0,4024 | -0,1914 |
| 4 | -0,2926 | -0,4374 | -0,7082 | -0,1921 |
| 5 | -0,8262 | -0,7450 | -0,8857 | -0,6708 |
| 14 | -0,0161 | -0,0283 | -0,1716 | 0,0953 |
| 19 | 0,2014 | 0,2564 | 0,1455 | 0,3577 |
| 23 | -0,5134 | -0,4976 | -0,6145 | -0,3951 |
| 24 | -0,6032 | -0,5063 | -0,6612 | -0,3513 |
| 28 | -0,9015 | -0,6471 | -0,8431 | -0,4630 |
| 31 | -0,7252 | -0,7835 | -0,9417 | -0,6941 |
| 32 | -0,1113 | 0,1334 | -0,0650 | 0,3559 |
| 39 | -0,4393 | -0,4922 | -0,6879 | -0,3459 |
| 40 | -0,5357 | -0,6210 | -0,7749 | -0,4701 |
| 41 | -0,5283 | -0,5691 | -0,7407 | -0,4087 |
| 42 | -0,0814 | -0,0922 | -0,1708 | -0,0059 |
| 44 | -0,6248 | -0,6964 | -0,9860 | -0,4630 |
| 46 | -0,7307 | -0,7429 | -0,9345 | -0,6517 |

**Tabela III** – Parâmetros Estruturais (a, b, c).

| *Item* | $a_i$ | $\overline{a}_i$ | *I. C. (95%)* | | $c_i$ | $\overline{c}_i$ | *I. C. (95%)* | | $b_i$ | $\overline{b}_i$ | *I. C. (95%)* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,44 | 1,37 | 1,16 | 1,55 | 0,14 | 0,13 | 0,07 | 0,19 | -0,76 | -0,79 | -0,95 | -0,67 |
| 2 | 1,36 | 1,40 | 1,17 | 1,62 | 0,18 | 0,20 | 0,17 | 0,24 | 0,05 | 0,06 | -0,04 | 0,17 |
| 3 | 1,30 | 1,34 | 1,12 | 1,55 | 0,21 | 0,22 | 0,14 | 0,30 | -0,88 | -0,85 | -1,05 | -0,70 |
| 4 | 1,15 | 1,02 | 0,71 | 1,37 | 0,18 | 0,18 | 0,15 | 0,20 | 2,04 | 2,07 | 1,83 | 2,40 |
| 5 | 1,93 | 1,90 | 1,57 | 2,17 | 0,21 | 0,19 | 0,14 | 0,24 | -1,11 | -1,09 | -1,30 | -0,98 |
| 6 | 1,89 | 1,73 | 1,41 | 2,01 | 0,23 | 0,21 | 0,18 | 0,25 | 0,23 | 0,22 | 0,14 | 0,31 |
| 7 | 0,59 | 0,58 | 0,49 | 0,67 | 0,16 | 0,15 | 0,05 | 0,28 | -1,58 | -1,63 | -1,93 | -1,31 |
| 8 | 0,74 | 0,70 | 0,59 | 0,82 | 0,22 | 0,23 | 0,11 | 0,36 | -0,91 | -0,86 | -1,19 | -0,53 |
| 9 | 1,64 | 1,58 | 1,16 | 2,01 | 0,16 | 0,16 | 0,15 | 0,18 | 1,90 | 2,02 | 1,85 | 2,32 |
| 10 | 2,25 | 2,13 | 1,71 | 2,50 | 0,19 | 0,19 | 0,16 | 0,21 | 0,44 | 0,48 | 0,41 | 0,56 |
| 11 | 2,04 | 1,73 | 1,48 | 1,99 | 0,20 | 0,17 | 0,12 | 0,22 | -0,28 | -0,34 | -0,45 | -0,25 |
| 12 | 0,73 | 0,73 | 0,55 | 0,94 | 0,19 | 0,19 | 0,14 | 0,24 | 1,37 | 1,41 | 1,25 | 1,64 |
| 13 | 0,93 | 0,95 | 0,80 | 1,09 | 0,19 | 0,20 | 0,12 | 0,27 | -0,46 | -0,43 | -0,60 | -0,26 |
| 14 | 0,75 | 0,74 | 0,58 | 0,87 | 0,16 | 0,15 | 0,09 | 0,22 | 0,35 | 0,35 | 0,16 | 0,55 |
| 15 | 4,28 | 3,32 | 2,77 | 3,93 | 0,22 | 0,20 | 0,16 | 0,25 | -0,56 | -0,62 | -0,72 | -0,54 |
| 16 | 1,57 | 1,78 | 1,43 | 2,10 | 0,19 | 0,20 | 0,17 | 0,22 | 0,81 | 0,80 | 0,71 | 0,92 |
| 17 | 1,76 | 1,72 | 1,42 | 2,01 | 0,17 | 0,15 | 0,13 | 0,18 | 0,87 | 0,87 | 0,78 | 1,01 |
| 18 | 0,78 | 0,72 | 0,57 | 0,86 | 0,22 | 0,20 | 0,13 | 0,26 | 0,55 | 0,48 | 0,29 | 0,68 |
| 19 | 1,76 | 1,80 | 1,37 | 2,11 | 0,18 | 0,19 | 0,16 | 0,21 | 0,97 | 1,03 | 0,92 | 1,21 |
| 20 | 3,00 | 2,45 | 1,73 | 3,04 | 0,22 | 0,21 | 0,19 | 0,22 | 1,41 | 1,45 | 1,34 | 1,65 |
| 21 | 1,26 | 1,22 | 1,02 | 1,39 | 0,21 | 0,21 | 0,16 | 0,26 | -0,00 | 0,04 | -0,07 | 0,14 |
| 22 | 0,63 | 0,70 | 0,54 | 0,86 | 0,25 | 0,33 | 0,22 | 0,41 | -0,23 | 0,09 | -0,25 | 0,38 |
| 23 | 1,13 | 1,20 | 1,01 | 1,37 | 0,16 | 0,19 | 0,12 | 0,26 | -0,87 | -0,79 | -0,96 | -0,64 |
| 24 | 0,76 | 0,73 | 0,62 | 0,87 | 0,23 | 0,24 | 0,09 | 0,40 | -1,58 | -1,55 | -1,96 | -1,09 |
| 25 | 0,65 | 0,76 | 0,57 | 0,97 | 0,21 | 0,24 | 0,20 | 0,27 | 1,82 | 1,92 | 1,72 | 2,24 |
| 26 | 0,63 | 0,56 | 0,46 | 0,68 | 0,18 | 0,12 | 0,06 | 0,19 | 1,05 | 0,91 | 0,70 | 1,16 |
| 27 | 1,83 | 1,88 | 1,58 | 2,22 | 0,24 | 0,25 | 0,18 | 0,32 | -0,79 | -0,78 | -0,93 | -0,67 |
| 28 | 0,56 | 0,68 | 0,54 | 0,82 | 0,19 | 0,23 | 0,16 | 0,29 | 0,17 | 0,35 | 0,14 | 0,57 |
| 29 | 1,65 | 1,45 | 1,21 | 1,67 | 0,19 | 0,14 | 0,05 | 0,23 | -1,07 | -1,19 | -1,39 | -1,05 |
| 30 | 0,66 | 0,66 | 0,55 | 0,78 | 0,23 | 0,24 | 0,09 | 0,42 | -1,54 | -1,66 | -2,12 | -1,19 |
| 31 | 1,80 | 1,83 | 1,49 | 2,11 | 0,17 | 0,22 | 0,18 | 0,26 | -0,55 | -0,51 | -0,66 | -0,42 |
| 32 | 1,73 | 1,41 | 1,14 | 1,70 | 0,22 | 0,25 | 0,08 | 0,44 | -2,09 | -2,11 | -2,49 | -1,82 |
| 33 | 1,67 | 1,62 | 1,33 | 1,81 | 0,17 | 0,16 | 0,12 | 0,21 | -0,30 | -0,34 | -0,44 | -0,25 |
| 34 | 1,78 | 1,68 | 1,43 | 1,94 | 0,22 | 0,17 | 0,08 | 0,26 | -0,97 | -1,08 | -1,24 | -0,94 |
| 35 | 1,90 | 1,98 | 1,64 | 2,31 | 0,22 | 0,18 | 0,07 | 0,31 | -1,38 | -1,42 | -1,62 | -1,27 |
| 36 | 0,74 | 0,82 | 0,67 | 0,95 | 0,14 | 0,19 | 0,11 | 0,27 | -0,46 | -0,36 | -0,56 | -0,17 |
| 37 | 1,14 | 1,10 | 0,91 | 1,26 | 0,21 | 0,18 | 0,13 | 0,23 | 0,16 | 0,13 | 0,02 | 0,25 |
| 38 | 1,27 | 1,44 | 1,18 | 1,66 | 0,22 | 0,25 | 0,21 | 0,28 | 0,30 | 0,39 | 0,30 | 0,49 |
| 39 | 0,73 | 0,69 | 0,56 | 0,85 | 0,20 | 0,17 | 0,11 | 0,22 | 0,76 | 0,70 | 0,52 | 0,88 |
| 40 | 1,65 | 1,65 | 1,37 | 1,98 | 0,19 | 0,28 | 0,14 | 0,41 | -1,82 | -1,83 | -2,10 | -1,61 |
| 41 | 0,70 | 0,70 | 0,60 | 0,80 | 0,18 | 0,20 | 0,07 | 0,36 | -1,65 | -1,66 | -2,01 | -1,28 |
| 42 | 1,32 | 1,21 | 1,01 | 1,38 | 0,21 | 0,16 | 0,12 | 0,20 | 0,23 | 0,18 | 0,06 | 0,28 |
| 43 | 1,06 | 0,95 | 0,79 | 1,11 | 0,26 | 0,22 | 0,15 | 0,28 | 0,10 | 0,01 | -0,15 | 0,17 |
| 44 | 1,27 | 1,16 | 0,77 | 1,61 | 0,16 | 0,15 | 0,13 | 0,17 | 1,77 | 1,91 | 1,69 | 2,28 |
| 45 | 1,62 | 1,64 | 1,36 | 1,89 | 0,22 | 0,23 | 0,18 | 0,28 | -0,34 | -0,32 | -0,43 | -0,21 |
| 46 | 1,98 | 1,92 | 1,61 | 2,20 | 0,16 | 0,18 | 0,13 | 0,23 | -0,92 | -0,96 | -1,15 | -0,84 |
| 47 | 0,56 | 0,53 | 0,44 | 0,62 | 0,21 | 0,16 | 0,07 | 0,27 | -0,26 | -0,42 | -0,73 | -0,07 |
| 48 | 1,18 | 1,08 | 0,90 | 1,25 | 0,21 | 0,22 | 0,11 | 0,34 | -0,95 | -0,98 | -1,22 | -0,75 |
| 49 | 0,70 | 0,70 | 0,58 | 0,83 | 0,18 | 0,21 | 0,11 | 0,30 | -0,22 | -0,19 | -0,46 | 0,06 |
| 50 | 0,92 | 0,82 | 0,63 | 1,00 | 0,18 | 0,19 | 0,15 | 0,22 | 1,20 | 1,25 | 1,10 | 1,46 |

As estimativas dos parâmetros correspondentes à estrutura de regressão explicativa para o DIF – parâmetros $\gamma_{02}$ e, $\gamma_{12}$, ver Tabela IV – foram muito similares às obtidas pelo ajuste de um modelo de regressão linear clássico, utilizando mínimos quadrados ordinários, a partir das estimativas encontradas para os parâmetros $d_{2i}^{b}$. Nesse caso, os resultados encontrados foram, respectivamente, 0,252 (0,90), 0,348 (0,131).

**Tabela IV** – Parâmetros da Estrutura de Regressão (Gama).

| *Grupo* | $\gamma_{0g}$ | $\overline{\gamma}_{0g}$ | *I.C. (95%)* | | $\gamma_{1g}$ | $\overline{\gamma}_{1g}$ | *I.C. (95%)* | | $\tau_g^2$ | $\overline{\tau}_g^2$ | *I.C. (95%)* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0,30 | 0,25 | 0,03 | 0,48 | 0,30 | 0,34 | 0,03 | 0,67 | 0,04 | 0,11 | 0,05 | 0,25 |

Um exemplo com 5 grupos foi construído de forma semelhante a esse com dois grupos e os resultados foram igualmente satisfatórios. A qualidade da estimação das proficiências, por exemplo, se manteve para todos os grupos focais.

Exemplo 2 (2 grupos – DIF na discriminação)

Nos exemplos anteriores foi considerada a possibilidade de DIF apenas nos parâmetros de dificuldade dos itens. Em geral, neste exemplo repetem-se as condições de simulação utilizadas no exemplo 2, porém o DIF é introduzido na discriminação dos modelos dos itens. Os itens para os quais se admite a possibilidade de um DIF foram escolhidos aleatoriamente através de uma distribuição *Bin*(50; 0,4), de tal forma que 31 dos itens foram escolhidos como itens âncoras e, portanto, os outros 19 itens foram considerados, então, como itens sujeitos a um possível funcionamento diferencial. Foram fixados os valores $\gamma_{02} = 0,5$ e, $\gamma_{12} = 0,2$, o que torna os itens com DIF menos discriminantes no grupo 2, e $W_{1i} = 1$ para 10 desses itens e $W_{1i} = 0$ para os demais, simulando o efeito de uma co-variável binária associada a uma possível característica comum desses dez itens. Os parâmetros $d_{2i}^{a}$ foram gerados conforme o modelo em (2) supondo que $\eta_{1i}^{b} \sim N(0;0,2)$ e $\tau_g^2 = 0,04$.

Da mesma forma que no exemplo anterior a convergência das cadeias de todos os parâmetros foi testada e através do critério RGR foram confirmadas.

As Tabelas de V a VII trazem os parâmetros gerados, as estimativas obtidas com o amostrador de GIBBS e os respectivos intervalos de credibilidade. Foram utilizadas as médias das 8000 últimas realizações das cadeias.

**Tabela V** – Parâmetros das Distribuições de Proficiências.

| **Média das proficiências dos grupos** | | | | | **Desvio Padrão das proficiências dos grupos** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Grupo** | **valor real** | **estimado** | **intervalo** | | **Grupo** | **valor real** | **estimado** | **intervalo** | |
| 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| 2 | 0,1500 | 0,1256 | 0,0668 | 0,1827 | 2 | 1,0000 | 0,9862 | 0,9335 | 1,0395 |

**Tabela VI** – Parâmetros Estruturais – DIF na discriminação.

| Parâmetro d | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Item** | **valor real** | **estimado** | **intervalo** | | **Item** | **valor real** | **estimado** | **intervalo** | |
| 2 | 0,3141 | 0,3914 | 0,1357 | 0,6613 | 35 | 0,6667 | 0,7162 | 0,5259 | 0,9342 |
| 4 | 0,5114 | 0,5623 | 0,3957 | 0,7559 | 37 | 0,6390 | 0,7113 | 0,5259 | 0,9319 |
| 8 | 0,8210 | 0,7715 | 0,5298 | 1,0316 | 41 | 0,4256 | 0,4203 | 0,2536 | 0,5779 |
| 13 | 0,2831 | 0,3953 | 0,1900 | 0,6091 | 42 | 0,8953 | 1,0447 | 0,8616 | 1,2405 |
| 15 | 0,2360 | 0,1856 | 0,0517 | 0,3327 | 43 | 0,6792 | 0,7679 | 0,5378 | 1,0209 |
| 18 | 0,5518 | 0,5026 | 0,2611 | 0,7330 | 44 | 0,5089 | 0,4419 | 0,2539 | 0,6386 |
| 21 | 0,9751 | 1,0194 | 0,8058 | 1,2436 | 45 | 0,4748 | 0,3923 | 0,2637 | 0,5326 |
| 23 | 0,6934 | 0,9308 | 0,7187 | 1,1738 | 47 | 0,2041 | 0,2606 | -0,0365 | 0,5237 |
| 24 | 0,8335 | 0,8903 | 0,6599 | 1,1376 | 48 | 0,8743 | 0,8891 | 0,6820 | 1,1183 |
| 30 | 0,6380 | 0,6288 | 0,4229 | 0,8620 | | | | | |

Em geral a estimação dos parâmetros a, b e c, foi tão satisfatória quanto no exemplo com DIF na dificuldade, por isso omitiu-se a apresentação dos resultados. Apenas no item 17 o modelo não conseguiu estimar bem o parâmetro a, ao se observar o gráfico desta cadeia (omitido), nota-se que não houve convergência. Provavelmente, a razão de não convergência esteja associada à dificuldade de estimação dos parâmetros desse item. Ele apresenta um nível de dificuldade muito baixo (trata-se, portanto, de um item muito fácil) e um nível de discriminação muito alto, isso é muito raro de acontecer na prática. A estimação dos parâmetros de DIF $d_{2i}^{a}$ também se mostrou bastante satisfatória, apenas o intervalo de credibilidade do item 23 não conteve o verdadeiro valor do parâmetro. No entanto, como era de se esperar, as estimativas foram um pouco piores do que no caso do DIF na dificuldade. A correlação de Pearson entre os parâmetros verdadeiros e as estimativas foi de 0,906.

**Tabela VII** – Parâmetros de Regressão – DIF na discriminação.

| Coeficientes da regressão do parâmetro d | | | | | Variância do erro da regressão | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Parâmetro** | **valor real** | **estimado** | **intervalo** | | **Grupo** | **valor real** | **estimado** | **intervalo** | |
| $\gamma_0$ | 0,5 | 0,4823 | 0,2774 | 0,6858 | 2 | 0,04 | 0,0820 | 0,0364 | 0,1696 |
| $\gamma_1$ | 0,2 | 0,2696 | -0,0069 | 0,5540 | | | | | |

A recuperação dos parâmetros correspondentes à estrutura de regressão explicativa para o DIF encontrado – parâmetros $\gamma_{02}$ e, $\gamma_{12}$ – foi também muito boa. A estimativa da variância de tal regressão não ficou tão boa quanto esperada, mas o intervalo de credibilidade conteve o verdadeiro valor do parâmetro. Finalmente, pode-se também dizer que a estimação dos parâmetros da distribuição das proficiências, média e desvio padrão do grupo focal, foi razoável, no entanto, não tão boa quanto no caso de DIF na dificuldade.

Os resultados indicam, portanto, que a presença de DIF expressivo na discriminação pode conduzir a maiores problemas na comparabilidade das médias dos dois grupos do que no caso de DIF na dificuldade. Porém, tendo em vista que na prática raramente se tem situações de DIF na discriminação nos níveis apresentados neste exemplo, garante-se, também, uma boa comparabilidade de resultados de testes educacionais aplicados a diferentes grupos de alunos.

Exemplo 3 (real)

Em Soares, Genovez & Galvão (2005) foi apresentada uma análise sistemática do funcionamento diferencial exibido pelos itens de geografia utilizados para avaliar os alunos da 4ª série do ensino fundamental no PROEB/SIMAVE-2001. O funcionamento dos itens foi analisado comparativamente para as diferentes regiões do estado de Minas Gerais, representadas pelos pólos regionais de ensino: pólo 1 – que compreende a região Metropolitana de Belo Horizonte; pólo 2 – região centro-sul; pólo 3 – região do triângulo mineiro; pólo 4 – região da Zona da Mata e pólo 5 – região Norte do estado. Naquele trabalho, três etapas distintas da investigação foram realizadas: na primeira os itens que exibiam funcionamento diferencial foram identificados com o uso do programa *Bilog-mg* utilizando-se modelos da teoria da Resposta ao Item para grupos múltiplos (*cf.* Bock & Zimovski, 1995); na segunda o funcionamento diferencial identificado foi confirmado ou não através das estatísticas clássicas fornecidas pelo SisAni e, a partir delas, o DIF encontrado foi classificado como desprezível, pequeno, intermediário ou grande; finalmente na terceira etapa a natureza do funcionamento diferencial foi investigada, sendo que conexões e extrapolações empíricas foram construídas de tal forma que se pôde chegar a uma interpretação sistêmica. Dezessete itens, dentro de um conjunto de 81 aplicados, exibiram algum funcionamento diferencial mais relevante. Sete desses itens avaliaram o conhecimento do aluno sobre as diferenças entre o espaço urbano e o espaço rural. Se ele seria capaz, por exemplo, de discriminar os produtos do campo de os produtos da cidade. Como, em geral, esses itens foram mais fáceis para os alunos das demais regiões do estado do que para os alunos do pólo Capital, pôde-se inferir que, provavelmente, o fato de os alunos dessas regiões estarem mais próximos de zonas rurais se refletiu numa maior competência nas respostas a essas questões. Infelizmente, as conclusões obtidas são genéricas, baseadas no bom senso e carecem de uma confirmação embasada em uma maior evidência estatística.

Utilizou-se, então, neste trabalho, o novo método apresentado para confirmar ou não as conclusões apresentadas em Soares *et al.* (2005). Lembrando, o teste analisado avaliou os alunos da 4ª série do ensino fundamental da rede estadual de ensino do Estado de Minas Gerais em 2001. Tomando como o grupo de referência a região metropolitana de Belo-Horizonte, e adotando-se como co-variável explicativa para o DIF na dificuldade uma variável dicotômica $W_{1i} = 1$ – para os itens associados à competência mencionada – e, $W_{1i} = 0$ para os demais, obteve-se os seguintes resultados após 20000 realizações das cadeias (São apresentadas as médias das últimas 10000 realizações).

**Tabela VIII** – Média e desvio-padrão dos Grupos.

| *Grupo* | $\overline{\mu}_g$ | *Intervalo de credibilidade* | | $\overline{\sigma}_g$ | *Intervalo de credibilidade* | |
|---|---|---|---|---|---|---|
| 1 – Metropolitana de BH | 0 | 0 | 0 | 1,0000 | 1,0000 | 1,0000 |
| 2 – Centro-Sul | 0,0938 | 0,0456 | 0,1450 | 1,0039 | 0,9626 | 1,0535 |
| 3 – Triângulo | 0,1224 | 0,0518 | 0,1945 | 0,9674 | 0,8987 | 1,0367 |
| 4 – Zona da Mata | 0,0801 | 0,0275 | 0,1336 | 1,0358 | 0,9819 | 1,0436 |
| 5 – Norte | -0,4178 | -0,4671 | -0,3673 | 1,0718 | 1,0253 | 1,1281 |

Os resultados apresentados na Tabela VIII são coerentes com os apresentados nos relatórios PROEB/SIMAVE-2001, indicando que os alunos da região norte do estado apresentam uma proficiência, em média, bastante inferior aos alunos das demais regiões na rede estadual de

ensino. No entanto, os resultados são diferentes dos obtidos sem consideração de DIF. Em particular, observou-se que as proficiências dos alunos do interior foram subestimadas, principalmente, dos alunos do triângulo mineiro.

De fato, quando se analisa as estimativas encontradas para os coeficientes da estrutura de regressão dos parâmetros de DIF, Tabela IX, observa-se que os itens que apresentaram DIF, em média, foram um pouco mais difíceis para a região do triângulo mineiro ( $\gamma_{03} = 0{,}1441$ ). Isto significa que o teste foi um pouco mais difícil para os alunos do triângulo mineiro.

**Tabela IX** – Coeficientes da Estrutura de Regressão do DIF (Gama).

| *Grupo* | $\overline{\gamma}_{0g}$ | *I.C. (95%)* | | $\overline{\gamma}_{1g}$ | *I.C. (95%)* | | $\overline{\tau}_g^2$ | *I.C.(95%)* | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0,0971 | -0,0225 | 0,2266 | -0,3638 | -0,5412 | -0,1953 | 0,0296 | 0,0152 | 0,0565 |
| 3 | 0,1441 | -0,0572 | 0,3477 | -0,3365 | -0,6128 | -0,0639 | 0,0677 | 0,0247 | 0,1532 |
| 4 | 0,0290 | -0,1596 | 0,2103 | -0,2060 | -0,4575 | 0,0590 | 0,0703 | 0,0326 | 0,1418 |
| 5 | 0,0786 | -0,0770 | 0,2308 | -0,2822 | -0,5065 | -0,0539 | 0,0435 | 0,0201 | 0,9100 |

As estimativas encontradas para $\gamma_{1g}$ , todas negativas nos grupos focais, indicam que de fato o conjunto de itens associados à competência mencionada é mais fácil para os alunos do interior do estado do que para os alunos da região metropolitana de Belo-Horizonte.

Além disso, o coeficiente $\overline{\gamma}_{1g}$ estimado permite ter uma idéia razoável do *gap* existente com relação a essa competência. No caso, os alunos da região metropolitana estão, algo em torno de 0,30 unidades do desvio padrão da distribuição de proficiências (dependendo da região), defasados em relação à competência exibida pelos alunos do interior do estado.

Finalmente, dos 17 itens originalmente diagnosticados com DIF através das estatísticas clássicas, apenas 13 deles foram confirmados pelo novo método.

## 6. Conclusões

Os resultados obtidos nos exemplos foram bastante promissores, sendo que para os modelos simulados a recuperação dos parâmetros gerados foi bastante satisfatória. Por outro lado, no exemplo real o resultado da análise confirmou a conclusão do estudo anterior de Soares, Genovez & Galvão (2005), baseada em estatísticas clássicas, de que itens associados à competência relacionada ao conhecimento do aluno sobre as diferenças entre o espaço urbano e o espaço rural são mais fáceis para alunos do interior do estado do que para a região metropolitana. Além disso, o modelo forneceu uma estimativa para o *gap* dessas diferenças (ver Tabela IX). Esses resultados encorajam a continuidade de estudos com o objetivo do aprimoramento do modelo e, conseqüente aplicação em outros estudos reais. No momento, estão sendo realizados exemplos que analisam as estimativas, simultaneamente, para o DIF na discriminação e na dificuldade do item. Além disso, estão sendo realizados estudos para avaliar o efeito do número de itens âncoras sobre as estimativas dos parâmetros e das proficiências. Estudos mais abrangentes do que o apresentado no exemplo 3, que incluam a estimação de $T_g$ , podem vir a ser muito interessantes, pois uma estimava dessa matriz de covariância permitiria prospectar co-variáveis associadas aos itens explicativas para o DIF

que ainda não tivessem sido incluídas no modelo. Por outro lado, como é possível construir uma distribuição estimada para essa matriz de covariância a partir das simulações obtidas pelo amostrador de Gibbs, poderia se testar hipóteses de correlação, por exemplo, para o DIF exibido pelos itens. A dificuldade aqui está associada à inclusão de um maior número de grupos para uma boa estimação de $T_g$. Alternativas podem ser adotadas, por exemplo, criando-se grupos artificiais e exigindo-se que os demais parâmetros dos modelos sejam iguais nesses grupos, estimando-se apenas $T_g$. Espera-se que esse artifício conduza a uma diminuição do tempo necessário de processamento, viabilizando-se uma análise da correlação apresentada pelos DIF dos diferentes itens.

**Agradecimentos**

Os autores agradecem ao CNPq e à FAPEMIG que apoiaram parcialmente este trabalho. Agradecem, ainda, ao CAEd pela seção dos dados utilizados. Agradecem, também, aos três revisores anônimos que muito contribuíram para a melhoria da última versão apresentada.

**Referências Bibliográficas**

(1) Baker, F.B. (1992). Item Response Theory. *STATISTICS*: Textbooks and Monographs n. 129. Marcel Dekker Inc, New York.

(2) Birnbaum, A. (1968). Some Latent Traits Models and Their Use in Inferring an Examinee's Ability. **In**: *Statistical Theories of Mental Test Scores* [edited by F. Lord and M. Novick], Addison-Wesley, Reading, MA, 397-472.

(3) Bock, D. & Zimovski, M.F. (1997). Multiple Group IRT. **In**: *Handbook of Modern Item Response Theory* [edited by R.W. Linden and R.K. Hambleton], Springer Verlag, New York, 433-448.

(4) Cole, N.S. (1993). History and Development of DIF. **In**: *Differential Item Functioning* [edited by P.W. Holland and H. Wainer], Lawrence Erlbaum, Hillsdale, NJ, 25-30.

(5) Dorans, N.J. & Holland, P.W. (1993). DIF detection and Description: Mantel-Haenszel and Standardization. **In**: *Differential Item Functioning* [edited by P.W. Holland and H. Wainer], Lawrence Erlbaum, Hillsdale, NJ, 35-66.

(6) Gamerman, D. & Lopes, H.L. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Second Edition. Taylor & Francis, New York.

(7) Hanson, B.A. (1998). Uniform DIF and DIF defined by Differences in Item Response Functions. *Journal of Educational and Behavioral Education*, **23**, 244-253.

(8) Longford, N.T.; Holland, P.W. & Thayer, D.T. (1993). Stability of the MH D-DIF Statistics Across Populations. **In**: *Differential Item Functioning* [edited by P.W. Holland and H. Wainer], Lawrence Erlbaum, Hillsdale, NJ, 171-196.

(9) Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum, Hillsdale, New Jersey.

(10) Patz, R.J. & Junker, B.W. (1999a). A Straightforward Approach to MCMC for Item Response Models. *Journal of Educational and Behavioral Statistics*, **24**, 146-178.

(11) Patz, R.J. & Junker, B.W. (1999b). Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses. *Journal of Educational and Behavioral Statistics*, **24**, 342-366.

(12) Soares, T.M.; Genovez, S.F. & Galvão, A.F. (2005). Análise do Comportamento Diferencial dos Itens de Geografia: estudo da 4ª série avaliada no PROEB/SIMAVE 2001. *Avaliação Educacional*, **16**, 81-110.

(13) Soares, T.M. & Galvão, A.F. (2004). *SISAni* – Um Sistema para a Análise de Itens: Manual do Usuário. UFJF, Caed, Faculdade de Educação (*mimeo*).

(14) Swaminathan, H. & Rogers, H.J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, **27**, 361-370.

(15) Swanson, D.B.; Brian, E.C.; Case, S.M.; Nungester, R.J. & Featherman, C. (2002). Analysis of Differential Item Functioning (DIF) Using Hierarchical Logistic Regression Models. *Journal of Educational and Behavioral Statistics*, **27**, 53-75.

(16) Thissen, D.; Steinberg, L. & Wainer, H. (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models. **In**: *Differential Item Functioning* [edited by P.W. Holland and H. Wainer], Lawrence Erlbaum, Hillsdale, NJ, 67-114.

## ANEXO A – Condicionais Completas

i) Proficiências.

Para gerar amostras de $p(\boldsymbol{\theta} \mid \boldsymbol{\beta}, \mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{T}, \mathbf{Y}, \mathbf{W})$ geram-se amostras, iterativamente, de:

$$p(\theta_j \mid \boldsymbol{\theta}_{\neq j}, \boldsymbol{\beta}, \mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{T}, \mathbf{Y}, \mathbf{W}) = p(\theta_j \mid \boldsymbol{\beta}_{I(j)}, \mathbf{d}_{g(j)I(j)}, \boldsymbol{\lambda}_{g(j)}, \mathbf{Y}_j) \propto$$
$$p(\mathbf{Y}_j \mid \theta_j, \boldsymbol{\beta}_{I(j)}, \mathbf{d}_{g(j)I(j)}, \boldsymbol{\lambda}_{g(j)})\ p(\theta_j \mid \boldsymbol{\beta}_{I(j)}, \mathbf{d}_{g(j)I(j)}, \boldsymbol{\lambda}_{g(j)}) =$$
$$= p(\mathbf{Y}_j \mid \theta_j, \boldsymbol{\beta}_{I(j)}, \mathbf{d}_{g(j)I(j)})\ p(\theta_j \mid \boldsymbol{\lambda}_{g(j)}) = \prod_{i \in I(j)} P(Y_{ij} \mid \theta_j, \beta_i, d_{g(j)i})\, p(\theta_j \mid \boldsymbol{\lambda}_{g(j)}),\ \ \forall\ j = 1, \ldots, J.$$

As igualdades se verificam, basicamente, admitindo-se que as respostas dos alunos são independentes, e as resposta atribuídas aos itens são independentes quando condicionadas às proficiências e aos parâmetros dos modelos dos itens. A distribuição acima não apresenta uma forma conhecida fechada, não permitindo que amostras sejam geradas diretamente. Por isso, vai-se empregar o algoritmo de *Metropolis-Hastings* com esse objetivo. Para tanto, adotou-se um núcleo de transição normal, tal que a proposição para o novo estado é gerada por:

$$\theta_j^{\ell} \sim q(\theta_j \mid \theta_j^{\ell-1}) = N(\theta_j^{\ell-1}, \sigma_\theta)$$

Adotou-se $\sigma_\theta = 0,2$, escolhida a partir de um estudo piloto de tal forma a garantir uma taxa adequada de transição da cadeia. As condições iniciais habituais para as proficiências adotadas foram os escores brutos padronizados obtidos por cada aluno no teste. Estudos de convergência para algumas proficiências foram realizados após a geração de cadeias em paralelo, com diferentes condições iniciais.

ii) Parâmetros das distribuições de proficiências dos grupos.

a) Média da distribuição de proficiências do grupo.

Admita que $p(\mu_g \mid .)$ representa distribuição a posteriori da média do grupo condicionada aos demais parâmetros do modelo. Nesse caso,

$$p(\mu_g \mid .) = p(\mu_g \mid \boldsymbol{\theta}_{J_g}, \sigma_g) \propto p(\boldsymbol{\theta}_{J_g} \mid \mu_g, \sigma_g)\ p(\mu_g \mid \sigma_g) \propto$$

$$\propto \prod_{i \in I(j)} p(\theta_j \mid \mu_g, \sigma_g)\ p(\mu_g \mid \sigma_g)\text{, e admitindo que } \mu_g \mid \sigma_g \sim N(0, \sigma_g)\text{, tem-se que:}$$

$$\mu_g \mid . \sim N(m_g, s_g),\ \ m_g = \frac{\sum_{j \in J_g} \theta_j}{N_g + 1} \text{ e, } \ s_g = \frac{\sigma_g}{\sqrt{N_g + 1}}.$$

$J_g$ representa o conjunto de alunos e $N_g$ o número de alunos do grupo g, $g = 1, ..., G$.

b) Variância da distribuição de proficiências do grupo.

No caso da variância das proficiências dos grupos:

$p(\sigma_g^2 \mid .) = P\left(\sigma_g^2 \mid \theta_{J_g}, \mu_g\right) \propto \prod_{i \in I(j)} p(\theta_j \mid \mu_g, \sigma_g^2)\, p(\sigma_g^2)$ e, adotando-se como prioris $\sigma_g^2 \sim GI(\alpha_g, \beta_g)$, onde $GI$ representa uma distribuição gama inversa, tem-se que:

$$\sigma_g^2 \mid . \sim GI\left(\alpha_g + \frac{N_g}{2}, \frac{\sum_{j \in J_g} \left(\theta_j - \mu_g\right)^2 + 2\beta_g}{2}\right),\ g = 1, ..., G.$$

iii) Parâmetros estruturais $\beta$.

Admitindo-se independência local dos itens, para se gerar amostras de $p(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{T}, \mathbf{Y}, \mathbf{W}) = p(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mathbf{d}, \mathbf{Y})$, pode-se gerar amostras a partir de:

$$p(\boldsymbol{\beta}_i \mid \boldsymbol{\theta}_{J(i)}, \mathbf{d}_i, Y_{J(i)}) \propto p(\mathbf{Y}_{J(i)} \mid \boldsymbol{\theta}_{J(i)}, \boldsymbol{\beta}_i, \mathbf{d}_i)\, p(\boldsymbol{\beta}_i \mid \mathbf{d}_i) =$$

$$= \prod_{j \in J(i)} P(Y_{ij} \mid \theta_j, \boldsymbol{\beta}_i, d_{ig(j)})\ p(\beta_i) = \prod_{j \in J(i)} P(Y_{ij} \mid \theta_j, \boldsymbol{\beta}_i, d_{ig(j)})\ p(a_i)\, p(b_i)\, p(c_i),\ \ \forall\ i = 1, ..., I.$$

Sendo que a última igualdade advém da hipótese de independência a priori dos parâmetros.

As prioris adotadas foram as seguintes: $a_i \sim LN(0, 2)$, $b_i \sim N(0, 1)$ e $c_i \sim beta(5, 17)$. Estas prioris são habitualmente empregadas, como por exemplo, são *defaults* no *software Bilog-mg*. E, novamente, como as distribuições apresentam formas desconhecidas o

algoritmo de Metropolis-Hastings foi empregado utilizando-se os seguintes núcleos de transição: $a_i^{\ell} \sim LN(\log(a_i^{\ell-1}), \alpha_a)$, $b_i^{\ell} \sim N(b_i^{\ell-1}, s_b)$ e, $c_i^{\ell} \sim U[c_i^{\ell-1} - \delta, c_i^{\ell-1} + \delta]$. Foram utilizados os seguintes valores para os parâmetros de dispersão dos núcleos de transição: $\alpha_a = 0{,}05,\ s_b = 0{,}2,\ \delta = 0{,}05.$

iv) Parâmetros estruturais de DIF.

Amostras de $p(\mathbf{d}^h \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{T}^h, \mathbf{Y}, \mathbf{W}^h, \mathbf{d}^{\neq h})$, $h = a, b$, podem ser obtidas a partir de amostras de:

$$p(d_{gi}^h \mid \mathbf{d}_{g,\neq i}^h, \mathbf{d}_g^{\neq h}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{T}^h, \mathbf{Y}, \mathbf{W}^h) = p(d_{gi}^h \mid \mathbf{d}_{g,\neq i}^h, \mathbf{d}_g^{\neq h}, \boldsymbol{\theta}_{J_g(i)}, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_g, \mathbf{T}^h, \mathbf{Y}_{J_g(i)}, \mathbf{W}^h) =$$

$$p(d_{gi}^h \mid \mathbf{d}_{g,\neq i}^h, \mathbf{d}_g^{\neq h}, \boldsymbol{\theta}_{J_g(i)}, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_g, \mathbf{T}^h, \mathbf{Y}_{J_g(i)}, \mathbf{W}^h) \propto$$

$$\propto p(Y_{J(i)} \mid \boldsymbol{\theta}_{J_g(i)}, \boldsymbol{\beta}_i, d_{gi}^h)\, p(d_{gi}^h \mid \mathbf{d}_{g,\neq i}^h, \mathbf{d}_g^{\neq h}, \mathbf{W}^h, \gamma_g, \mathbf{T}^h) =$$

$$= \prod_{j \in J_g(i)} p(Y_{ij} \mid \theta_j, \boldsymbol{\beta}_i, d_{gi})\, p(d_{gi}^h \mid \mathbf{W}_i^h, \boldsymbol{\gamma}_g^h, \tau_g^h),\ \ \forall\, i \in I_{dif}^h,\ (g = 2,...G),$$

Na última igualdade, admite-se que $\mathbf{T}^h = \left(\tau_g^h\right)^2 \mathbf{I}$, sendo $\mathbf{I}$ a matriz identidade de ordem $nid_h \times nid_h$, onde $nid_h$ o número de itens com provável DIF. Nesse caso, a priori condicional, adotada para os parâmetros de DIF foi a seguinte: $d_{gi}^h \mid \mathbf{W}_i^h, \boldsymbol{\gamma}_g^h, \tau_g^h \sim N(\mathbf{W}_i^h \boldsymbol{\gamma}_g^h, \left(\tau_g^h\right)^2)$. Já o núcleo de transição considerado foi o seguinte: $d_{gi}^{h,\ell+1} \sim N(d_{gi}^{h,\ell}, 0, 3)\ \forall\, i$. Por outro lado, no caso mais geral:

$$p(\mathbf{d}_g^h \mid \mathbf{d}_{\neq g}^h, \mathbf{d}_{\neq g}^{\neq h}, \boldsymbol{\theta}_{J_g}, \boldsymbol{\beta}, \boldsymbol{\gamma}_g, \mathbf{T}, \mathbf{Y}_{Jg}, \mathbf{W}) \propto p(\mathbf{Y}_{J(g)} \mid \boldsymbol{\theta}_{J_g}, \boldsymbol{\beta}, \mathbf{d}_g)\, p(\mathbf{d}_g^h \mid \mathbf{W}^h, \boldsymbol{\gamma}_g^h, \mathbf{T}^h) =$$

$$= \prod_{i \in I_{dif}^h} \prod_{j \in J_g} p(Y_{ij} \mid \theta_j, \boldsymbol{\beta}_i, d_{gi})\, p(\mathbf{d}_g^b \mid \mathbf{W}^h, \boldsymbol{\gamma}_g^h, \mathbf{T}^h),\quad (g = 2,...G).$$

sendo que $\mathbf{d}_g^h \mid \mathbf{W}^h, \boldsymbol{\gamma}_g^h, \mathbf{T}^h \sim N(\mathbf{W}^h \boldsymbol{\gamma}_g^h, \mathbf{T}^h)$. O núcleo de transição adotado foi o seguinte $\mathbf{d}_g^{h,\ell+1} \sim N(\mathbf{d}_g^{h,\ell}, 0, 2\mathbf{I})$.

v) Parâmetros da estrutura de regressão explicativa do DIF.

Para os parâmetros $\gamma$, devem ser geradas amostras de:

$$p(\boldsymbol{\gamma}_g^h \mid \boldsymbol{\gamma}_{\neq g}^h, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{T}_g^h, \mathbf{Y}, \mathbf{W}^h) = p(\boldsymbol{\gamma}_g \mid \mathbf{d}_g^b, \mathbf{T}_g^h, \mathbf{W}^h) \propto p(\mathbf{d}_g^b \mid \boldsymbol{\gamma}_g^h, \mathbf{W}^h, \mathbf{T}_g^h)\, p(\boldsymbol{\gamma}_g^h),\ \ g = 2,...,G,$$

Adotando-se a priori $\boldsymbol{\gamma}_g^h \sim N(\boldsymbol{\gamma}_0^h, \mathbf{S}_0^h)$, tem-se a distribuição condicional completa:

$$\boldsymbol{\gamma}_g^h \mid \mathbf{d}_g^h, \mathbf{T}_g^h, \mathbf{W}^h \sim N(\mathbf{H}, \mathbf{L}), \quad \mathbf{L} = \left[ \left(\mathbf{W}^h\right)^T \left(\mathbf{T}_g^h\right)^{-1} \mathbf{W}^h + \left(\mathbf{S}_0^h\right)^{-1} \right]^{-1} \text{ e,}$$

$$\mathbf{H} = \mathbf{L}\left[ \left(\mathbf{W}^h\right)^T \left(\mathbf{T}_g^h\right)^{-1} \mathbf{d}_g^h + \left(\mathbf{S}_0^h\right)^{-1} \boldsymbol{\gamma}_0^h \right].$$

vi) Duas situações foram consideradas para a estrutura de covariância $T$:

a) $\mathbf{T}_g^h = \left(\tau_g^h\right)^2 \mathbf{I}$,

Nesse caso, amostras de $p(\mathbf{T} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \mathbf{Y}, \mathbf{W})$ serão obtidas através de amostras de:

$$p(\left(\tau_g^h\right)^2 \mid \mathbf{d}_g^h, \boldsymbol{\gamma}_g^h, \mathbf{W}^h) \propto p(\mathbf{d}_g^h \mid \left(\tau_g^h\right)^2, \boldsymbol{\gamma}_g^h, \mathbf{W}^h)\, p(\left(\tau_g^h\right)^2) =$$

$$= GI(\alpha_g + \frac{nid}{2}, \left[ \frac{1}{2} \left(\mathbf{d}_g^h - \mathbf{W}^h\, \boldsymbol{\gamma}_g\right)^T \left(\mathbf{d}_g^h - \mathbf{W}^h\, \boldsymbol{\gamma}_g\right) + \boldsymbol{\beta}_g \right], \qquad g = 2, ..., G.$$

Onde $\left(\tau_g^h\right)^2 \sim GI(\alpha_g, \beta_g)$ a priori.

b) $\mathbf{T}_g^h = \mathbf{T}^h, \; g = 2, ..., G.$

Nesse caso, adotando-se como priori $\mathbf{T}^h \sim WI(\frac{nid_h}{2}, \mathbf{T}_0^h)$, onde $WI$ representa a distribuição de Wishart inversa, tem-se que:

$$\mathbf{T}^h \mid \mathbf{d}^h, \boldsymbol{\gamma}^h, \mathbf{W}^h \sim WI(\frac{nid_h + G}{2}, \frac{1}{2}\sum_{g=1}^{G} \left(\mathbf{d}_g^h - \mathbf{W}^h \boldsymbol{\gamma}_g^h\right)\left(\mathbf{d}_g^h - \mathbf{W}^h \boldsymbol{\gamma}_g^h\right)^T + \mathbf{T}_0^h).$$