# ESTIMAÇÃO BAYESIANA E MCMC EM MODELO MULTIDIMENSIONAL DA TRI

**Patrícia Costa**
Universidade da Beira Interior - Departamento de Matemática - Leacom
Rua Marquês d'Ávila e Bolama, 6200-001 Covilhã - Portugal
patriciamotacosta@gmail.com

**Tufi Soares**
Centro de Políticas Públicas e Avaliação da Educação - CAEd
Rua Eugênio Nascimento, Bairro Dom Orione, Juiz de Fora - MG, CEP: 36038-330
tufi@caed.ufjf.br

**Maria Eugénia Ferrão**
Universidade da Beira Interior - Departamento de Matemática
Rua Marquês d'Ávila e Bolama, 6200-001 Covilhã - Portugal
meferrao@ubi.pt

**Pedro Nuno Oliveira**
Universidade do Minho - Departamento de Produção e Sistemas
Campus de Azurém, 4800-058 Guimarães - Portugal
pno@dps.um.pt

**RESUMO**:
Neste trabalho pretende-se obter as estimativas dos parâmetros dos itens e dos factores latentes do modelo da Teoria de Resposta ao Item multidimensional logístico de dois parâmetros conjugando a estimação bayesiana com o uso de métodos de simulação Markov Chain Monte Carlo. Em particular, usa-se o algoritmo de Metropolis-Hastings com passos de Gibbs. Todas as etapas do algoritmo e respectiva fundamentação matemática apresentam-se e ilustram-se com recurso a computação desenvolvida em Matlab. Para testar o algoritmo proposto utilizam-se dados simulados, considerando que o factor latente afere 2 e 3 dimensões. Usa-se o critério de informação AIC para identificar o número de factores que melhor se ajusta aos dados. Para comparar as estimativas dos parâmetros obtidas pela aplicação do modelo com os valores verdadeiros utilizam-se as estatísticas: correlação de Pearson, Erro Absoluto Médio e Erro Quadrático Médio.

**Palavras-chave:** Modelo multidimensional da Teoria da Resposta ao Item, Markov Chain Monte Carlo, Estimação de parâmetros, Simulação.

**ABSTRACT**:
The aim of this work is to obtain item and latent factors parameter estimates of the multidimensional logistic two parameter model of Item Response Theory using bayesian estimation and simulation methods Markov Chain Monte Carlo. In particular, we use Metropolis-Hastings algorithm with steps of Gibbs. All steps of the algorithm and its mathematical fundament are presented and illustrated using a computer application developed in Matlab. To test the proposed algorithm we used simulated data, considering that the latent factor measure 2 and 3 dimensions. The AIC information criterion is used to identify the number of factors that better fit to the data. To compare the parameter estimates obtained by applying the model with the true values the statistics used are: Pearson correlation and Mean Absolute Error Mean Square Error.

**Keywords:** Multidimensional model of Item Response Theory, Markov Chain Monte Carlo, Parameter estimation, Simulation.

# 1 Introdução

Os modelos da Teoria de Resposta ao Item (TRI) relacionam formas de representar a probabilidade de um examinando responder correctamente a um item tendo em conta os seus factores latentes/habilidades, na área de conhecimento avaliada, e as características/propriedades do item.

Para itens dicotómicos e considerando que existe apenas uma população envolvida, existem instrumentos em que os examinandos devem possuir mais do que um factor latente para responderem a um item ou em que os itens requerem mais do que um factor latente para serem respondidos correctamente. Assim, é necessário recorrer a modelos que consideram múltiplos parâmetros para os examinandos e que são designados por modelos multidimensionais (Reckase, 2009). Esta classe de modelos da TRI é a mais complexa e permite avaliar o comportamento dos examinandos, isto é a resposta ao item, dados os factores latentes dos examinandos e os parâmetros que representam as características do item. Nos modelos multidimensionais da TRI (MMTRI) as respostas aos itens estão associadas a múltiplos factores latentes e a sua utilização é importante para reflectir o nível de factor latente/habilidade nas diferentes dimensões para cada examinando.

A utilização de MMTRI iniciou-se a partir dos anos 70, baseada nos trabalhos de Paul Horst (1965), Roderick McDonald (1967), Bengt Muthén (1978), Fumiko Samejima (1974), entre outros. A história e o desenvolvimento dos MMRI baseiam-se principalmente em duas áreas: a análise factorial (Horst, 1965; Christoffersson, 1975; Muthén, 1978; McDonald, 1977 e 1985; Bock e Aitkin, 1981 e Bejar, 1977 e a extensão dos modelos de resposta ao item unidimensionais (Rasch, 1962; Lord e Novick, 1968 e Samejima, 1974. Estas duas áreas distinguem-se pela metodologia aplicada e pela forma como os resultados são interpretados.

A estimação em modelos da TRI envolve dois tipos de parâmetros, os factores latentes dos examinandos e os parâmetros dos itens. Os procedimentos de estimação de parâmetros em modelos da TRI mais utilizados são máxima verosimilhança e bayesianos. As limitações na utilização de procedimentos de máxima verosimilhança são: não estão definidos para alguns padrões de resposta, como itens com acerto total ou erro total e respostas omissas, a possibilidade das estimativas dos parâmetros dos itens não pertençam ao intervalo esperado, a ocorrência frequente de problemas numéricos em testes longos, e a dificuldade na aplicação do algoritmo EM no caso dos modelos multidimensionais. O uso de procedimentos de estimação bayesianos em modelos da TRI item exige cálculos mais complexos do que os procedimentos de máxima verosimilhança. Para colmatar as limitações dos procedimentos de estimação de máxima verosimilhança e bayesianos e a necessidade de utilizar um procedimento de estimação que permita a sua extensão à medida que aumenta a complexidade dos modelos, fez com que se conjugasse a estimação bayesiana com o uso de métodos de simulação de *Markov Chain Monte Carlo* (MCMC). A utilização destes métodos permite uma diminuição das dificuldades na estimação dos modelos da TRI a nível teórico e a nível computacional.

Segundo Gilks, Richardson e Spieglhalter (1996) os métodos MCMC oferecem muito potencial para a estimação dos parâmetros dos modelos da TRI. Conceptualmente, a criação de um algoritmo de MCMC para um modelo da TRI é semelhante ao algoritmo EM, mas como MCMC não envolve quadratura numérica exacta (para o passo E) ou o pré-cálculo das derivadas (para o passo M), então MCMC é mais fácil de implementar. O método MCMC é usado para gerar as amostras da distribuição a posteriori conjunta dos parâmetros dos modelos da TRI.

O primeiro autor a usar MCMC em modelos da TRI foi Albert (1992), que obteve as estimativas dos parâmetros do modelo de 2 parâmetros de ogiva normal para dados aumentados usando *Gibbs Sampling*. Patz e Junker (1999a, 1999b) usaram MCMC em modelos unidimensionais da TRI que consideram respostas omissas e em modelos para itens politómicos. Em 2001, Béguin e

Glas (2001) generalizaram o procedimento exposto por Albert (1992) para estimar o modelo de 2 parâmetros da ogiva normal. Estes autores utilizaram um procedimento bayesiano, através da aplicação do método MCMC para estimar o modelo de 3 parâmetros da ogiva normal e expuseram uma generalização do procedimento ao modelo multidimensional para o factor latente. O procedimento permite analisar dados para múltiplas populações e *designs* incompletos. Num *design* incompleto apenas um subconjunto do conjunto total de itens é administrado a cada examinando. No entanto, é assumido que a matriz de covariância subjacente para identificar os factores latentes é uma matriz identidade, o que não é realista, uma vez que os factores latentes aferidos num teste, em geral, são correlacionados. De la Torre e Patz (2002) propuseram a estimação simultânea dos factores latentes de modelos multidimensionais usando MCMC. No entanto, neste trabalho assume-se que os parâmetros dos itens são conhecidos, o que na prática não costuma acontecer. Em 2003, Bolt e Lall (2003) investigaram a estimação dos parâmetros dos itens de modelos multidimensionais compensatórios e não compensatórios usando o algoritmo *Metropolis-Hastings* em MCMC. A principal limitação apontada para este trabalho é de ser considerado apenas o modelo bidimensional.

Neste trabalho, propõe-se um procedimento de estimação bayesiano, através do uso de MCMC, para estimar os parâmetros dos itens e dos factores latentes do modelo multidimensional da TRI compensatório logístico de 2 parâmetros, recorrendo ao algoritmo de *Metropolis-Hastings* com passos de *Gibbs*. A estrutura do trabalho é a que se descreve seguidamente. Na secção 2, apresenta-se a metodologia adoptada baseada em modelos da TRI. Nesse sentido, especifica-se formalmente o modelo multidimensional logístico de 2 parâmetros e descrevem-se os procedimentos de estimação utilizados. Na secção 3 apresentam-se os resultados das análises com dados simulados, obtidos pela aplicação do algoritmo MCMC proposto ao modelo. Começa-se por apresentar os resultados considerando que os dados aferem 2 factores latentes. Seguidamente ilustram-se os resultados obtidos no caso em que a dimensão do factor latente é 3. Para ambos os casos, descreve-se como são gerados os dados e comparam-se os resultados obtidos para as simulações. Na última secção, apresentam-se as principais conclusões.

# 2 Metodologia

## 2.1 Modelo logístico multidimensional de 2 parâmetros da TRI. Especificação formal do modelo

O modelo multidimensional logístico de 2 parâmetros é um modelo compensatório que se baseia numa combinação linear das coordenadas dos factores latentes. Este modelo é uma extensão do modelo unidimensional de 2 parâmetros da TRI (Birnbaum, 1968; Hambleton, Swaminathan e Rogers, 1991) e é dado por:

$$P(U_{ij} = 1|\theta_j, a_i, d_i) = \frac{e^{a_i\theta'_j + d_i}}{1 + e^{a_i\theta'_j + d_i}}. \tag{1}$$

onde $i = 1, 2, ..., n$ e $j = 1, 2, ..., m$, com:
$k$ é o número de dimensões do factor latente;
$U_{ij}$ representa a resposta dicotómica do examinando $j$ ao item $i$, em que $U_{ij} = 1$, se for resposta correcta e $U_{ij} = 0$, se for resposta incorrecta;
$\theta_j$ é um vector $(1 \times k)$ dos $k$ factores latentes do $j$-ésimo examinando;
$P(U_{ij} = 1|\theta_j, a_i, d_i)$ é a probabilidade de um examinando $j$ com factores latentes $\theta_j$ responder correctamente ao item $i$;
$a_i$ é um vector $(1 \times k)$ referente aos parâmetros de discriminação do item $i$;

$d_i$ é um escalar que representa o parâmetro de dificuldade do item $i$;
$e$ é a base dos logaritmos neperianos (cujo valor é aproximadamente 2,72);
$n$ é o número de itens do teste;
$m$ é o número de examinandos.

O expoente de $e$ no modelo (1) pode ser escrito da seguinte forma:

$$a_i\theta'_j + d_i = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + ... + a_{ik}\theta_{jk} + d_i = \sum_{l=1}^{k} a_{il}\theta_{jl} + d_i. \quad (2)$$

O expoente é uma função linear de elementos de $\theta$ com o parâmetro $d$, como a ordenada na origem e os elementos do vector $a$ como os parâmetros de inclinação/discriminação. Uma das propriedades deste modelo é que a expressão representada no expoente define uma recta num espaço de $k$ dimensões.

## 2.2 MCMC em TRI

MCMC consiste em gerar amostras da distribuição de interesse a partir de distribuições que constituam uma cadeia de Markov. Estas distribuições são as chamadas distribuições de transição da cadeia e devem ser adequadamente escolhidas de forma a que a cadeia convirja em distribuição para a distribuição de interesse que, no caso de modelos da TRI, corresponde à distribuição a posteriori conjunta dos parâmetros do modelo. Após se atingir a convergência, as amostras são geradas a partir dessa distribuição estacionária. O objectivo é gerar uma amostra com uma dimensão suficientemente grande, a partir desta distribuição estacionária, que aproxime bem a distribuição a posteriori conjunta exacta. Para isso, estipula-se um *burn-in* que é um número de iterações que se julgue serem necessárias para a convergência da cadeia, e, a partir daí, gera-se uma amostra suficientemente grande para se obter uma boa aproximação da distribuição de interesse.

Existem alguns métodos para a construção das cadeias, entre eles o *Gibbs Sampling* (Geman e Geman, 1984) e que são casos especiais de uma estrutura geral desenvolvida por Metropolis *et al.* (1953) e Hastings (1970), que é o algoritmo designado *Metropolis-Hastings* (MH).

No caso do *Gibbs Sampling*, foi introduzido por Geman e Geman (1984) e com grande contribuição de Gelfand e Smith (1990). O método consiste em considerar as distribuições condicionais completas dos parâmetros como as distribuições de transição da cadeia de Markov. Assume-se que a distribuição a posteriori conjunta de interesse seja $\pi(\theta, X)$, com $\theta = (\theta_1, \theta_2, ..., \theta_k)'$, onde cada compomente de $\theta$ pode ser um escalar, um vector ou uma matriz. Dessa forma, as distribuições condicionais completas são $\pi_l(\theta_l) = \pi_l(\theta_{-l})$, $l = 1, 2, ..., k$ onde $\theta_{-l}$ é o vector $\theta$ sem a $l$-ésima componente.

É de notar que uma condição necessária do *Gibbs Sampling* é que as distribuições condicionais completas sejam conhecidas e que se saiba gerar valores dessas, porém, isso nem sempre é possível. Uma solução para este problema é o algoritmo *Metropolis-Hastings*. Neste caso, suponha-se que se deseja gerar uma amostra de uma distribuição de interesse através de cadeias de Markov, que no caso é a distribuição à posteriori, e defina-se $p(\theta, \phi)$ como a distribuição de transição que leva à convergência da cadeia para a distribuição de interesse. O algoritmo MH consiste em gerar os valores de uma distribuição de transição arbitrária com densidade $q(\theta, \phi)$, baseado numa probabilidade $\alpha(\theta, \phi)$ tal que:

$$p(\theta, \phi) = q(\theta, \phi)\alpha(\theta, \phi), \text{ se } \theta \neq \phi.$$

Assim, $q(\theta, \phi)$ define uma densidade para $p(\theta, .)$ para todo o valor possível do parâmetro diferente de $\theta$. Consequentemente, existe uma probabilidade restante da cadeia permanecer em $\theta$ dada por:

$$p(\theta,\theta) = 1 - \int q(\theta,\phi)\alpha(\theta,\phi)d\phi$$

Baseado em Hastings (1970), a expressão considerada para a probabilidade de aceitação é

$$\alpha(\theta,\phi) = min\left\{1, \frac{\pi(\phi)q(\phi,\theta)}{\pi(\theta)q(\theta,\phi)}\right\}$$

Em modelos da TRI, em geral, são conhecidas apenas algumas distribuições condicionais completas. Assim, neste trabalho é adoptado o método conhecido como *Gibbs Sampling* com passos *Metropolis* proposto por Muller (1991). As componentes que não podem ser geradas das amostras directamente a partir das respectivas condicionais completas $\pi_i$, são geradas as amostras de $\pi_i$ através de uma sub-cadeia de *Metropolis-Hastings* dentro do ciclo amostral de *Gibbs*. Isto é, essas componentes permitem gerar uma amostra da proposta $q_i$ e aceites com probabilidade $\alpha(\theta,\phi)$. Quanto mais parecidas forem a proposta $q_i$ e a condicional completa $\pi_i$, mais próxima de 1 será a probabilidade de aceitação.

Seguidamente propõe-se o procedimento para se fazer inferência sobre o modelo logístico multidimensional de 2 parâmetros, utilizando esta abordagem. Para isso, começa-se por apresentar as seguintes especificações: função de verosimilhança, distribuições à priori, distribuição à posteriori conjunta, distribuições condicionais completas e o algoritmo de aplicação de MCMC.

**Função de verosimilhança**

A função de verosimilhança (Baker e Kim, 2004) é dada por:

$$L(U|\Theta,A,d) = \prod_{j=1}^{m}\prod_{i=1}^{n} p_i(\theta_j)^{U_{ij}}(1-p_i(\theta_j))^{(1-U_{ij})}$$

onde,
$p_i(\theta_j)$ representa a probabilidade do examinando $j$ responder correctamente ao item $i$, dada pelo modelo multidimensional logístico de 2 parâmetros, isto é, $p_i(\theta_j) = P(U_{ij} = 1|\theta_j, a_i, d_i)$;
$U$ é a matriz de respostas dos $m$ examinandos aos $n$ itens;
$\Theta$ é uma matriz $m \times k$ que representa todos os factores latentes aferidos no instrumento;
$A$ é uma matriz $n \times k$ que representa todos os parâmetros $a$ para os $n$ itens;
$d$ é o vector do parâmetro de dificuldade para os $1 \times n$ itens.

**Distribuições à priori**

Assume-se que os parâmetros são independentes a priori, o que significa que a priori conjunta (Baker e Kim, 2004) é dada por:

$$\pi(\Theta,A,d) = \pi_\Theta(\Theta)\pi_A(A)\pi_d(d)$$

ou

$$\pi(\Theta,A,d) = \sum_{j=1}^{m}\pi_\Theta(\Theta)\sum_{i=1}^{n}\pi_A(A)\pi_d(d)$$

Para cada componente, especificam-se seguintes distribuições à priori:

- $a_i \sim NM(\mu_a, \Omega_a)$ onde $NM(.,.)$ representa a distribuição normal multivariada com vector média $\mu_a$ e matriz de covariâncias $\Omega_a$ ($\mu_a$ é um vector com $k$ zeros e $\Omega_a$ é a matriz identidade $I_k$);

- $b_i \sim N(\mu_b, c_b^2)$, onde $N(.,.)$ representa a distribuição normal com $\mu_b = 0$ e $c_b^2 = 2$;

- $\theta_j \sim NM(\mu_\theta, \Omega_\theta)$ onde $\mu_\theta$ é o vector nulo com $k$ componentes e $\Omega_\theta$ é a matriz identidade $I_k$;

**Distribuição à posteriori conjunta**

A distribuição à posteriori conjunta é utilizada para obter as estimativas de todos os parâmetros do modelo e foi obtida pela aplicação do teorema de Bayes. Assume-se que as distribuições à priori para os itens e os factores latentes são independentes. A distribuição à posteriori conjunta (Baker e Kim, 2004) é dada por:

$$p(A, d, \Theta|U) \propto L(U|\Theta, A, d)\pi(\Theta, A, d)$$

Em geral, é difícil obter a constante de normalização para a expressão anterior, o que significa que não é possível encontrar uma forma analítica fechada para a distribuição à posteriori.

**Distribuições condicionais completas**

A distribuição condicional completa (Baker e Kim, 2004) para os parâmetros referentes aos factores latentes dos examinandos, pela aplicação do teorema de Bayes, é

$$P_\theta(\theta_j|\Theta_{-j}, A, d, U) = \prod_{i=1}^{n} p_i(\theta_j)^{U_{ij}}(1 - p_i(\theta_j))^{(1-U_{ij})}\pi_\theta(\theta_j)$$

onde, $\Theta_{-j} = (\theta_1, \theta_2, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_m)'$.

De modo análogo se obtêm as distribuições condicionais completas para os parâmetros dos itens. Isto é,

$$P_a(a_i|A_{-i}, \Theta, d, U) = \prod_{j=1}^{m} p_i(\theta_j)^{U_{ij}}(1 - p_i(\theta_j))^{(1-U_{ij})}\pi_a(a_i)$$

e

$$P_d(d_i|d_{-i}, \Theta, A, U) = \prod_{j=1}^{m} p_i(\theta_j)^{U_{ij}}(1 - p_i(\theta_j))^{(1-U_{ij})}\pi_d(d_i)$$

onde,
$A_{-i}$ é uma matriz $(n-1) \times k$, isto é, $A_{-i} = (a_1, a_2, ..., a_{i-1}, a_{i+1}, ..., a_n)$;
$d_{-i}$ é um vector com $(n-1)$ componentes, isto é, $d_{-i} = (d_1, d_2, ..., d_{i-1}, d_{i+1}, ..., d_n)$.

**Algoritmo**

O algoritmo da aplicação de MCMC ao modelo multidimensional logístico de 2 parâmetros foi implementado em Matlab e apresenta-se seguidamente.

Os valores iniciais adoptados para os parâmetros de discriminação dos itens são $a_i^0 = 1$; para o parâmetro de dificuldade dos itens são $d_i^0 = 0$ e para todos os factores latentes são $\theta_j^0 \sim N(0, 1)$.

O algoritmo utilizado é então o seguinte:
1) Inicializa-se o contador de iterações da cadeia $s = 1$;
2) Gera-se, $\forall j = 1, ..., m$, os factores latentes $\theta_j^s$. Para obter $\theta_j^s$, a partir de $p(\theta|A^{s-1}, d^{s-1}, U)$:

a) Gera-se $\theta_j^* \sim NM(\mu_{\theta_j}|\mu_{\theta_j^{s-1}}, \Omega_{\theta_j})$ independentemente para cada $j = 1, ..., m$;

b) Calculam-se os vectores de $m$ probabilidades de aceitação, $\alpha(\theta_j^{s-1}, \theta_j^*) = min\{1, R_\theta\}$ onde $r_\theta = \frac{L(U_j|\theta_j^*, A^{s-1}, d^{s-1})\pi(\theta_j^*)}{L(U_j|\theta_j^{s-1}, A^{s-1}, d^{s-1})\pi(\theta_j^{s-1})}$ para $j = 1, ..., m$;

c) Aceita-se $\theta_j^s = \theta_j^*$ com probabilidade $\alpha_j$, caso contrário tem-se $\theta_j^s = \theta_j^{s-1}$.

3) Gera-se $\forall i = 1, ..., n$, os parâmetros $d_i^s$. Para obter $d^s$ a partir de $p(d|\theta^{s-1}, A_{s-1}, U)$:

a) Gera-se $d_i^* \sim N(d_i|d_i^{s-1}, c_{d_i}^2)$ independentemente para cada $i = 1, ..., n$ (Em particular foi considerado $c_{d_i}^2 = 0,01$);

b) Calcula-se o vector de $n$ probabilidades de aceitação $\alpha(d_i^{s-1}, d_i^*) = min\{1, R_d\}$, onde $r_d = \frac{L(U_i|d_i^*,\theta^{s-1},A^{s-1})\pi(d_i^*)}{L(U_i|d_i^{s-1},\theta^{s-1},A^{s-1})\pi(d_i^{s-1})}$ para $i = 1, ..., n$;

c) Aceita-se $d_i^s = d_i^*$ cada com probabilidade $\alpha_i$, caso contrário faz-se $d_i^s = d_i^{s-1}$.

4) Gera-se $\forall i = 1, ..., n$, os parâmetros $a_i^s$. Para obter $a^s$ a partir de $p(A|\theta^{s-1}, d^{s-1}, U)$:

a) Gera-se $a_i^* \sim NM(\mu_{a_i}|\mu_{a_i^{s-1}}, \Omega_{a_i})$ independentemente para cada $i = 1, ..., n$;

b) Calculam-se os vectores de $n$ probabilidades de aceitação, $\alpha(a_i^{s-1}, a_i^*) = min\{1, R_a\}$, onde $R_a = \frac{L(U_i|a_i^*,\theta^{s-1},d^{s-1})\pi(a_i^*)}{L(U_i|a_i^{s-1},\theta^{s-1},d^{s-1})\pi(a_i^{s-1})}$ para $i = 1, ..., n$;

c) Aceita-se $a_i^s = a_i^*$ com probabilidade $\alpha_i$, caso contrário tem-se $a_i^s = a_i^{s-1}$.

5) Muda-se o contador de $s$ para $s + 1$ e retorna-se ao passo 2 até que seja gerado o tamanho da amostra desejado (considerando um *burn-in*).

## 3 Dados e Resultados

O procedimento de estimação proposto foi aplicado a dois conjuntos de dados simulados. Foi considerado que o factor latente aferia 2 e 3 dimensões. Em particular, os dados referem-se ás respostas de 2000 examinandos a 40 itens e ás respostas de 2500 examinandos a 80 itens, respectivamente.

**Caso 1: Duas dimensões para os factores latentes**

Foram simuladas as respostas de 2000 examinandos a um instrumento constituído por 40 itens dicotómicos. Para a obtenção das respostas dos examinandos aos itens geraram-se aleatoriamente os valores verdadeiros de $a$, $d$ e $\theta$, considerando 2 factores latentes. O parâmetro de dificuldade foi gerado a partir de uma distribuição normal, com média zero e desvio-padrão 1,4. Para o parâmetro $a$ foram considerados, para os dois factores, os valores de 0 e 1, em 20 itens, 1 e 0 em 10 itens, e, 1 e 1 para os restantes itens. Os factores latentes foram gerados a partir de uma distribuição normal multivariada, com vector média zero e com covariância entre os dois factores 0,3. Foi calculada a probabilidade de resposta correcta de cada examinando a cada item, pela aplicação do modelo multidimensional logístico de 2 parâmetros e foram geradas as respostas dicotómicas a partir da distribuição binomial. A base de dados correspondente foi inserida no programa em Matlab, a partir do qual foram obtidas as estimativas dos parâmetros dos itens $a$ e $d$ e as estimativas dos factores latentes considerando um número diferente de factores. A distribuições à priori utilizadas foram as definidas anteriormente. No algoritmo proposto, para cada iteração, foram obtidas as estimativas dos factores latentes de cada examinando, seguidamente foi estimado o parâmetro de dificuldade de cada item e no final foram obtidas as estimativas dos parâmetros de discriminação de cada item em cada dimensão do factor latente. Para obter as estimativas dos parâmetros do modelo, factores latentes e parâmetros dos itens, foi gerado o parâmetro independentemente para cada

examinando ou item, foram calculadas as respectivas probabilidades de aceitação e foi aceite ou rejeitado o novo valor proposto de cada parâmetro considerando o valor da probabilidade obtido anteriormente. Em particular, foram geradas 15000 amostras, descartando as 14000 primeiras para *burn-in*. As estimativas do factor latente e dos parâmetros dos itens foram obtidas através do cálculo da média dos valores de cada parâmetro nas últimas 1000 iterações.

Para a seleção dos modelos em termos do número de dimensões dos factores latentes foi utilizado o critério de informação *Akaike information criterion* (AIC) (Akaike, 1981) considerando 1, 2 e 3 factores. O critério AIC é dado por

$$AIC = -2log(L(U|\Theta, A, d)) + 2K \tag{3}$$

onde,
$K$ é o número de parâmetros estimados pela aplicação do modelo.

A função de verosimilhança e os restantes parâmetros já foram definidos anteriormente.

Os valores obtidos para este critério constam na tabela **??**.

Tabela 1: AIC para dados simulados

| Número de factores | AIC |
|---|---|
| 1 | $1,327 \times 10^5$ |
| 2 | $7,774 \times 10^4$ |
| 3 | $8,103 \times 10^4$ |

Com o recurso ao critério AIC, um modelo é mais apropriado quanto menor for essa estatística. Assim, pela análise da tabela **??**, o número de factores que deve ser considerado e que proporciona um melhor ajuste dos dados é dois.

Para 2 factores latentes, foram comparados os valores verdadeiros com as estimativas obtidas para parâmetros do modelo através do cálculo da correlação de Pearson (tabela **??**).

Tabela 2: Correlação entre os valores verdadeiros e as estimativas obtidas

| | Correlação |
|---|---|
| Corr($a_1$,$\hat{a}_1$) | 0,976 |
| Corr($a_2$,$\hat{a}_2$) | 0,917 |
| Corr($d$,$\hat{d}$) | 0,994 |
| Corr($\theta_1$,$\hat{\theta}_1$) | 0,777 |
| Corr($\theta_2$,$\hat{\theta}_2$) | 0,880 |

A correlação entre os valores verdadeiros dos parâmetros de discriminação e de dificuldade dos itens e as estimativas obtidas varia entre 0,92 e 0,99. No que se refere à comparação entre valores verdadeiros dos dois factores latentes e as estimativas obtidas, a correlação varia entre 0,78 e 0,88.

Adicionalmente, foram calculadas as estatísticas Erro Absoluto Médio (EAM) e o Erro Quadrático Médio (EQM), com vista a comparar as estimativas obtidas para os parâmetros dos itens e para os factores latentes, com os valores verdadeiros de cada parâmetro. Os valores obtidos para estas estatísticas são apresentados na tabela **??**.

A análise da tabela permite constatar que, para todos os parâmetros do modelo, o EQM apresenta valores inferiores comparativamente com o EAM. Em particular, o EAM varia entre $0,08$ e $0,509$ e o EQM apresenta valores compreendidos entre $0,013$ e $0,41$. Adicionalmente, para ambas as estatísticas os valores mais elevados são os que correspondem à comparação das estimativas dos dois factores latentes com os valores verdadeiros.

Tabela 3: Estatísticas EAM e EQM

| | EAM | EQM |
|---|---|---|
| $(a_1,\hat{a}_1)$ | 0,080 | 0,013 |
| $(a_2,\hat{a}_2)$ | 0,177 | 0,046 |
| $(d,\hat{d})$ | 0,136 | 0,035 |
| $(\theta_1,\hat{\theta}_1)$ | 0,509 | 0,410 |
| $(\theta_2,\hat{\theta}_2)$ | 0,375 | 0,223 |

Os tempos de execução do programa em Matlab, num computador Intel core 2, 2 gb de RAM e processador t7200 a 2,00 Ghz, são apresentados na tabela **??**.

Tabela 4: Tempo de execução do programa

| Número de factores | Tempo (em minutos (m) e segundos (s)) |
|---|---|
| 1 | 32m 38s |
| 2 | 33m 31s |
| 3 | 34m 53s |

Uma das críticas feitas à aplicação de algoritmos MCMC é o tempo de execução dos programas, que dependem, da própria eficiência do programa, do número de examinandos e de itens, da velocidade de convergência e das características do computador utilizado. Nesta aplicação verifica-se que, comparativamente com outros softwares como WINBUGS (Bolt e Lall, 2003) e SPLUS (Patz e Junker, 1999a e 1999b) o tempo de execução é inferior aos demais, apesar das características dos computadores utilizados serem diferentes.

**Caso 2: Três dimensões para os factores latentes**

A base de dados simulada é constituída pelas respostas de 2500 examinandos a um instrumento constituído por 80 itens dicotómicos. Neste caso, foram considerados 3 factores latentes. Para a obtenção das respostas dos examinandos aos itens geraram-se aleatoriamente os valores verdadeiros de $a$, $d$ e $\theta$. Foram consideradas para o parâmetro $a$, as sequências de valores apresentadas na tabela **??**.

Tabela 5: Parâmetros de discriminação verdadeiros

| Sequência | Número de itens |
|---|---|
| 011 | 16 |
| 010 | 24 |
| 111 | 12 |
| 110 | 8 |
| 101 | 12 |
| 100 | 8 |

Os parâmetros dos itens e os factores latentes foram tal como foi definido no caso de duas dimensões para os factores latentes. Foram geradas 15000 amostras para as simulações e as estimativas finais foram obtidas pelo cálculo da média das estimativas de cada parâmetro nas últimas 1000 iterações.

Para apurar o número de factores que melhor se adequa aos dados foi calculado o critério de informação AIC (tabela **??**).

A análise da tabela anterior permite concluir que o critério AIC apresenta valor inferior para o caso em que o número de factores é três. Assim, pode-se concluir que devem ser considerados 3 factores latentes para a análise dos dados.

Com o propósito de comparar as estimativas dos parâmetros de discriminação e dificuldade dos itens e dos factores latentes com os valores verdadeiros foi calculada a correlação de Pearson.

Tabela 6: AIC para dados simulados

| Número de factores | AIC |
|---|---|
| 1 | $8,351 \times 10^5$ |
| 2 | $1,738 \times 10^5$ |
| 3 | $1,718 \times 10^5$ |
| 4 | $1,746 \times 10^5$ |

Os resultados obtidos são apresentados na tabela **??**.

Tabela 7: Correlação entre os valores verdadeiros e as estimativas obtidas

| | Correlação |
|---|---|
| Corr($a_1$,$\hat{a}_1$) | 0,753 |
| Corr($a_2$,$\hat{a}_2$) | 0,426 |
| Corr($a_3$,$\hat{a}_3$) | 0,750 |
| Corr($d$,$\hat{d}$) | 0,7235 |
| Corr($\theta_1$,$\hat{\theta}_1$) | 0,835 |
| Corr($\theta_2$,$\hat{\theta}_2$) | 0,568 |
| Corr($\theta_3$,$\hat{\theta}_3$) | 0,743 |

Em geral, verifica-se que a correlação entre os valores verdadeiros e as estimativas obtidas para os parâmetros dos itens é positiva forte, com excepção do segundo factor. O mesmo se verifica para a correlação entre as estimativas dos factores latentes e os valores verdadeiros destes parâmetros.

As estatísticas Erro Absoluto Médio (EAM) e o Erro Quadrático Médio (EQM) são apresentadas na tabela **??**.

Tabela 8: Estatísticas EAM e EQM

| | EAM | EQM |
|---|---|---|
| ($a_1$,$\hat{a}_1$) | 0,247 | 0,097 |
| ($a_2$,$\hat{a}_2$) | 0,409 | 0,312 |
| ($a_3$,$\hat{a}_3$) | 0,385 | 0,236 |
| ($d$,$\hat{d}$) | 1,161 | 3,509 |
| ($\theta_1$,$\hat{\theta}_1$) | 0,431 | 0,298 |
| ($\theta_2$,$\hat{\theta}_2$) | 0,686 | 0,744 |
| ($\theta_3$,$\hat{\theta}_3$) | 0,550 | 0,485 |

Em geral, o EQM apresenta valores inferiores comparativamente com o EAM, com excepção do parâmetro de dificuldade. Os parâmetros de discriminação dos itens apresentam o EAM e o EQM mais baixos e o parâmetro de dificuldade dos itens apresenta o maior erro comparativamente com o respectivo valor verdadeiro.

Os tempos de execução do programa em Matlab, considerando as respostas de 2500 examinandos a 80 itens, e variando o número de factores latentes são apresentados na tabela **??**. As características do computador utilizado são: Intel core 2, 2 gb de RAM e processador t7200 a 2,00 Ghz.

Tal como seria de esperar, quanto mais factores foram considerados maior é o tempo de execução do programa. No entanto, considerando diferentes números de factores verifica-se que o tempo de execução do programa variou no máximo 5 minutos.

Tabela 9: Tempo de execução do programa

| Número de factores | Tempo (em horas (h) e minutos (m) |
|---|---|
| 1 | 1h 54m |
| 2 | 1h 56m |
| 3 | 1h 58m |
| 4 | 1h 59m |

# 4 Discussão

Neste trabalho utilizaram-se procedimentos de simulação MCMC para estimar os parâmetros do modelo multidimensional compensatório logístico de 2 parâmetros. O algorimto que se propõe permite obter as estimativas dos parâmetros dos itens e dos factores latentes dos examinandos simultaneamente. Para isso, usou-se o algoritmo de Metropolis-Hastings com passos de Gibbs. Recorre-se a dados simulados para testar o procedimento de estimação proposto, considerando que os dados aferem 2 e 3 factores latentes. Os dados referem-se à resposta de 2000 examinandos a 40 itens e à resposta de 2500 examinandos a 80 itens, respectivamente. Utilizou-se o critério de informação AIC com vista a identificar o número de factores que melhor se ajustava aos dados. Definido o número de factores, para comparar as estimativas dos parâmetros obtidas pela aplicação do modelo com os valores verdadeiros de cada parâmetro, calculá-mos: a correlação de Pearson, o Erro Absoluto Médio e o Erro Quadrático Médio.

A análise dos resultados de ambas as simulações permitiram-nos concluir que, obtêm-se boas estimativas para todos os parâmetros do modelo compensatório multidimensional logístico de 2 parâmetros, e que a metodologia adoptada para fazer inferência é eficiente para estimar os parâmetros do modelo. Os valores iniciais escolhidos foram obtidos através da realização de simulações. O algoritmo proposto permite estimar simultaneamente os parâmetros associados aos itens e aos factores latentes dos examinandos e pode ser generalizado a outros modelos. Em particular, para a generalização deste modelo ao multidimensional logístico de 3 parâmetros, bastará apenas considerar o parâmetro de acerto ao acaso como parâmetro do item. Este parâmetro é geral para cada item, e como tal a sua estimação pode ser efectuada como se tratasse de um modelo unidimensional. Outra das vantagens da utilização desta abordagem é a possibilidade da sua aplicação a dados reais. Para a utilização da abordagem proposta ainda podemos apontar as seguintes vantagens: a possibilidade da generalização do algoritmo a outros modelos multidimensionais, por exemplo os modelos não compensatórios e a aplicação desta abordagem a dados que aferem um maior número de factores latentes.

# Referências

[1] **Akaike, H.** (1981), Likelihood of o model and information criteria, *Journal of Econometrics* 16, 3-14.

[2] **Albert, J.H.** (1992), Bayesian estimation of normal ogive item response curves using Gibbs sampling, *Journal of Educational Statistics*, 17, 251-269.

[3] **Andrade, D.F., Tavares, H. R. e Valle, R.C.**, *Teoria da Resposta ao Item: conceitos e aplicações*. 14$^o$ Sinape - Associação Brasileira de Estatística, Caxambu, 2000.

[4] **Baker, F.B. e Kim, S.H.**, *Item Response Theory - Parameter Estimation Techniques* - 2nd ed., Marcel Dekker Inc, New York, 2004.

[5] **Béguin, A.A. e Glas, C.A.W.** (2001), MCMC Estimation and Model-fit analysis of Multidimensional IRT Models, *Psychometrika*, 66, 541-562.

[6] **Bejar, I. I.** (1977), An application of the continuous response level model to personality measurement, *Applied Psychological Measurement*, 1, 509-521.

[7] **Birnbaum, A.**, Some latent trait models and their use in inferring an examinee's ability, In Lord, F. e Novick M. (Eds.), *Statistical Theories of Mental Test Scores*, Addison-Wesley, New York, 1968.

[8] **Bock, R.D. e Aitkin M.** (1981), Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm, *Psychometrika*, 46, n. 4, 443-459.

[9] **Bolt, D.M. e Lall, V.F.** (2003), Estimation of Compensatory and Noncompensatory Multidimensional Item Response Models using Markov Chain Monte Carlo, *Applied Psychological Measurement*, 27, n. 6, 395-414.

[10] **Christoffersson, A.** (1975), Factor analysis of dichotomized variables, *Psychometrika*, 40, 5-32.

[11] **Gelfand, A.E. e Smith, A.F.M.** (1990), Sampling-based approaches to calculating marginal densities, *Journal of American Statistical Association*, 85, 398-409.

[12] **Geman, S., Geman, D.** (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

[13] **Gilks, W.R., Richardson, S. e Spiegelhalter, D.J.**, Introducing Markov chain Monte Carlo, In Gilks, W.R., Richardson, S. e Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, Washington DC, 1-19, 1996.

[14] **Hambleton, R.K., Swaminathan, H. e Rogers, H.J.**, *Fundamentals of Item Response Theory*, Sage Publications, North Caroline, 1991.

[15] **Hastings, W.K.** (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97-109.

[16] **Horst, P.**, *Factor analysis of data matrices*, Holt, Rinehart & Winston, New York, 1965.

[17] **Lord, F.M. e Novick, M.R.**, *Statistical Theories of Mental Test Scores*, Addison-Wesley, New York, 1968.

[18] **McDonald, R.P.** (1967), Nonlinear factor analysis, *Psychometric Monograph*, n. 15, 1-167.

[19] **McDonald, R.P.**, *Factor analysis and related methods*, Erlbaum, Hillsdale NJ, 1985.

[20] **Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.** (1953), Equation of State Calculations by Fast Computing Machine, *Journal of Chemical Physics*, 21, 1087-1091.

[21] **Muller, P.** (1991), Metropolis based posterior integration schemes, *Technical Report*, Statistics Department, Purdue University.

[22] **Muthén, B.** (1978), Contributions to factor analysis of dichotomous variables, *Psychometrika*, 43, 551-560.

[23] **Patz, R.J. e Junker, B.W.**(1999a), A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models, *Journal of Educational and Behavioral Statistics*, 24, n. 2, 146-178.

[24] **Patz, R. J. e 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, n. 4, 342-366.

[25] **Rasch, G.** (1962), On general laws and the meaning of measurement in psychology, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4, 321-334.

[26] **Reckase, M.D.** (2009), *Mutidimensional Item Response Theory*, Springer, New York.

[27] **Samejima, F.** (1974), Normal ogive model on the continuous response level in the multidimensional space, *Psychometrika*, 39, 111-121.

[28] **De la Torre, J. e Patz, R.J.** (2002), *A Multidimensional Item Response Theory Approach to Simultaneous Ability Estimation*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, United States of America: New Orleans.