

**MÉTODOS DE ESTIMAÇÃO  
NA TEORIA DE RESPOSTA AO ITEM**

**Caio Lucidius Naberezny Azevedo**

DISSERTAÇÃO APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO GRAU DE MESTRE  
EM  
ESTATÍSTICA

Área de Concentração: **Estatística**  
Orientador: **Prof. Dr. Dalton Francisco de Andrade**

*Durante a elaboração deste trabalho o autor recebeu apoio financeiro do CNPq*

– São Paulo, Fevereiro de 2003 –

**MÉTODOS DE ESTIMAÇÃO  
NA TEORIA DE RESPOSTA AO ITEM**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Caio Lucidius Naberezny Azvedo e aprovada pela comissão julgadora.

São Paulo, 27 de Fevereiro de 2003.

Comissão Julgadora:

- Prof. Dr. Dalton Francisco de Andrade (orientador) - INE/UFSC
- Prof. Dr. Heleno Bolfarine - IME/USP
- Prof. Dr. Ruben Klein - LNCC/CNPQ

Dedico esta parte de minha  
vida e essência

À Deus acima de tudo que existe, existiu e existirá pela  
dádiva da vida

À Jesus pela redenção e lição de amor

À meus pais, Verônica e Hélio, pela personalidade, caráter e  
pelos carinhos incomensuráveis

À minha namorada Fabiene, pela ternura, os sonhos e a des-  
coberta

À meus mestres, com entusiasmo, respeito e gratidão, Prof.  
Maurício e Prof. Dalton.

Para aqueles que acreditam nos sonhos...

*“Ó mar salgado, quanto do teu sal  
São lágrimas de Portugal!  
Por te cruzarmos, quantas mães choraram,  
Quantos filhos em vão rezaram!  
Quantas noivas ficaram por casar  
Para que fosses nosso, ó mar!  
Valeu a pena?  
Tudo vale a pena  
Se a alma não é pequena.  
Quem quiser passar além do Bojador  
Tem que passar além da dor.  
Deus ao mar o perigo e o abismo deu,  
Mas nele é que espelhou o céu.”*

Fernando Pessoa

Para aqueles que amam como se jamais pudessem se machucar...

Metal Contra as Nuvens : (Renato Russo)

I

Não sou escravo de ninguém. Ninguém senhor do meu domínio. Sei o que devo defender, e por valor eu tenho e temo o que agora se desfaz.

Viajamos sete léguas. Por entre abismos e florestas. Por Deus nunca me vi tão só. É a própria fé o que destrói. Estes são dias desleais.

Sou metal - raio, relâmpago e trovão Sou metal, eu sou o ouro em seu brasão. Sou metal: me sabe o sopro do dragão.

Reconheço o meu pesar: Quando tudo é traição, O que venho encontrar. É a virtude em outras mãos.

Minha terra é a terra que é minha, e sempre será minha terra. Tem a lua, tem estrelas e sempre terá.

II

Quase acreditei na sua promessa, e o que vejo é fome e destruição. Perdi a minha sela e a minha espada. Perdi o meu castelo e minha princesa.

Quase acreditei, quase acreditei.

E, por honra, se existir verdade. Existem os tolos e existe o ladrão. Há quem se alimente do que é roubo.

Vou guardar o meu tesouro. Caso você esteja mentindo.

Olha o sopro do dragão.

III

É a verdade o que assombra. O descaso que condena, a estupidez o que destrói.

Eu vejo tudo que se foi e o que não existe mais. Tenho os sentidos já dormentes, o corpo quer, a alma entende.

Esta é a terra-de-ninguém, e sei que devo resistir- Eu quero a espada em minhas mãos.

Sou metal - raio, relâmpago e trovão. Sou metal: eu sou o ouro em seu brasão. Sou metal: me sabe o sopro do dragão.

Não me entrego sem lutar - Tenho ainda coração. Não aprendi a me render: Que caia o inimigo então.

IV

- Tudo passa, tudo passará.

E nossa história não estará pelo avesso assim, sem final feliz. Teremos coisas bonitas pra contar.

E até lá, vamos viver. Temos muito ainda por fazer. Não olhe para trás - Apenas começamos.

O mundo começa agora - Apenas começamos.

Para os puros de coração...

“Na verdade, na verdade vos digo que aquele  
que crê em mim também fará as obras que eu faço,  
e as fará maiores do que estas,  
porque eu vou para meu Pai.  
E tudo quanto pedirdes em meu nome eu o farei,  
para que o Pai seja glorificado no Filho.  
Se pedirdes alguma coisa em meu nome, eu o farei.  
Se me amais, guardai os meus mandamentos.”

João 14, 12:15

“(And I) ride the winds of a brand new day  
High where mountain’s stand  
Found my hope and pride again  
Rebirth of a man”

Rafael Bittencourt

“Where the streets have no name  
We’re still building  
Then burning down love  
Burning down love  
And when I go there  
I go there with you  
It’s all I can do”

Bono Vox

*Para Fabiene, com amor e ternura...*

“You’re in my mind all of the  
time I know that’s not enough  
If the sky can crack  
there must be some way back  
For love and only love  
Electrical Storm  
Electrical Storm  
Baby don’t cry”

Bono Vox

“Oh darling, hear my soul and heed my cry  
Cause all my crying  
may flood an ocean in my heart”

André Matos

“Teu corpo alimenta meu espírito  
Teu espírito alegra minha mente  
Tua mente descansa meu corpo  
Teu corpo aceita o meu como a um irmão  
Longe, longe, estou em outra estação”

Renato Russo



---

## Agradecimentos

---

O prezado leitor pode se sentir à vontade, para que, em um primeiro momento, exclua essa parte de nosso trabalho e ir adiante, pois a não-leitura do mesmo não irá comprometer a compreensão de nossos desenvolvimentos. No entanto advertimos que essa é uma parte mais do que imprescindível, pois nela externamos os nossos mais sinceros e profundos agradecimentos.

Gostaríamos, assim, de agradecer

À Deus, pois pela imensidão de amor que representa, criou e cuida de tudo que existe de bom e puro. Por nos dar continuamente inúmeras chances de estarmos em comunhão com Ele e de desfrutarmos a sua chama de amor eterna. Mais ainda, numa demonstração de ternura, permitiu que caminhássemos e aprendêssemos com seu filho unigênito Jesus e assim nos redimir de todo o pecado.

À Jesus, por seus ensinamentos e sabedoria, lições de fé, esperança e amor, sendo este último o mais importante. Tudo que nos mostrou, desde dar a outra face e amar nossos inimigos até a acreditar no impossível e ter a certeza da vitória. Convocou-nos para sermos plena e verdadeiramente santos e nos ensinou, acima de tudo, o que o amor é capaz de fazer.

À meu orientador Prof. Dalton (“mestre Jedi Quin-Dalt-Jan”), por toda a nossa mais que agradável convivência, desde a graduação, por nossa relação ir além da mais sincera amizade, pelo espelho que representa para mim. Sem dúvida um excelente exemplo de profissional, pai, amigo e professor, pelo qual também me sinto como filho. Por todos os conselhos, acadêmicos e de vida, pelo apoio e a crença em nossa capacidade. As dúvidas tiradas dentro de fora da sala. Sim, muito aprendi com o senhor, mais do que esperava e menos do que merecia. Esperamos que possamos sempre trabalhar juntos e cultivar essa relação por toda a eternidade.

Ao nosso mais que amado mestre e professor Maurício (“mestre Jedi Obi-Mau-Kenobi”) por todas as nossas conversas que sempre vão do âmbito acadêmico até as experiências intermináveis dessa vida. Por ter me ajudado a moldar meu caráter e construir minha personalidade. Sempre disposto a estender a mão não importando a gravidade da situação. Por representar a gênese de minha vida como estatístico. Por representar um pai, um ponto de referência e apoio do qual sempre posso esperar os mais gratuitos conselhos e votos de confiança. Vejo que todas as palavras sobre os caminhos da vida mostraram-se mais do que verdadeiras e valiosas. Senão agreguei de modo pleno em minha vida tudo que recebi do senhor, ao menos sincretizei parte em minha essência.

Ao Prof. Júlio da Motta Singer, do IME-USP, pela recepção mais do que paterna, por me

ajudar a tornar mais agradável a minha permanência em São Paulo, os conselhos, a solicitude, os ensinamentos, os trabalhos, enfim, por tudo o que me ajudou a amadurecer enquanto ser humano e profissional. Uma das poucas pessoas que conheço que conseguem unir de modo *sui generis* caráter e conhecimento.

Aos meus pais Verônica e Hélio, por tudo que recebi durante minha vida, os carinhos, as broncas e os conselhos. Principalmente minha mãe, Verônica (“dulcíssima madre”) pelo seu sacrifício para conceder a mim e a meus irmãos todas as possibilidades para nos tornarmos pessoas lúcidas, felizes e capazes de conseguirmos nosso próprio futuro. Por ter me dado todas as chaves do universo, e mais ainda, por ter me ensinado a conseguir a sabedoria necessária para escolher quais portas abrir. Minha adorável mãezinha, essa é a nossa vitória, por todos os anos de amor, carinho e ensinamentos.

Aos meus irmãos André e Elisângela, pelo apoio e as brincadeiras de infância (que ainda perduram até hoje). Principalmente ao meu irmão André por toda a ajuda e o carinho e por sempre acreditar em nosso sucesso.

A minha namorada Fabiene (“Ene”), pelos sonhos, o encontro de “Arcádia”, os “Carpe Diem”, as ternuras e as poesias. A percepção de que o tempo é apenas uma invenção, de que tudo que emana da alma é infinito, de que jamais seremos livres enquanto não transpusermos as “Muralhas Eternas” e nos lançarmos ao mar da esperança em busca de terras sagradas. Por me ajudar a acreditar que é possível viver juntos para sempre “onde as ruas não tem nome”. Por compartilhar de modo único desde as felicidades mais imperceptíveis até as vitórias inimagináveis. Sobretudo, por me ensinar que nada floresce se não for poética e amorosamente cultivado.

A todos os professores e funcionários do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, por toda a minha formação durante a Graduação, os bate-papos descontraídos, o apoio para a conclusão da Graduação e do Mestrado, e por toda ajuda a mim dispensada. Dentro os quais gostaria de destacar : Maurício, Welliandre, Robson, André, Ana Maria, Sílvia (professores) Márgeri, Luiza, Mariluze e Margarida (funcionários).

Aos componentes da minha banca, Prof. Heleno e Prof. Ruben Klein, pelas sugestões e comentários mais do que valiosos para o aperfeiçoamento de nossa dissertação.

A todos os professores e funcionários do Departamento de Estatística dos IME-USP, pela formação durante o Mestrado, as amizades construídas e o apoio imprescindível. Dentre os professores gostaria de destacar Rinaldo, Antônio Carlos, Beth, Márcia, Cléia, Gilberto e Heleno, pela receptividade e pela amizade construída. Em especial a Cecília, um verdadeiro oásis de solicitude e candura, por toda a atenção e ajuda durante o Mestrado.

Aos meus amigos da Graduação, pelas brincadeiras e apoio, pelas piadas (ainda que sem graça) e por fazerem parte da minha vida. Em especial a Agnaldo (“cabra da peste”), Andrea Mendes, Eveline, Olavo, Márcio, Iésus, Evelyny (“neném”) e Patrícia (“depende”).

A meus amigos Ronald e Gilberto, pelas discussões sobre a TRI, a amizade, o apoio e incentivos constantes, principalmente durante o Mestrado.

Ao mais que irmão e camarada Juvêncio, pelas brincadeiras no CRUSP e no IME-USP, pelas discussões probabilísticas e estatísticas, pelas idas ao rodízio, a amizade mais do que sincera, o apoio nos momentos de tristeza, os risos nos momentos de alegria e a lição de que não é preciso sair do mesmo ventre para cultivar o sentimento de fraternidade.

A todos os amigos de Fortaleza-CE, que ficaram sempre na mais "fiel torcida": César, Rômulo, Yuri, Fabiano, Júnior, Léo e Daniel, por todos os momentos de risos e festa e também de conversas sérias, as peladas e tudo mais, e a Mayara por todo o apoio.

Aos meus amigos "cruspianos" Alex e Álvaro, por ajudarem ao tempo passar nos momentos de ócio criativo, as conversas, as músicas e principalmente pela experiência de vida que me proporcionaram e ao aprendizado.

A todos os meus amigos da Pós-Graduação do IME-USP, pelo convívio, amizade, brincadeiras (principalmente na Sala 250) e pelo aprendizado. Dentro os quais destaco: Adrilayne ("tacacá e açaf"), Edvaldo ("ei moço"), Elias ("Chaibub"), Alberto ("chaid"), Marcelo ("charon"), Marcelo ("carioca"), Anderson ("picareta"), Gisela ("Sandy"), Eriê ("manda por e-mail"), Diana ("Shakira"), Marcos e Daniela ("Presidente Prudente"), Ângela, Fábio ("candango"), Michel, Jorge ("Pata") e Jose ("Colo-colo").

E a todos que não mencionei que me ajudaram de uma forma ou de outra a cumprir essa missão.

---

## Resumo

---

Neste trabalho apresentamos os mais importantes processos de estimação em algumas classes de modelos de resposta ao item (Dicotômicos e Policotômicos). Discutimos algumas propriedades desses métodos. Com o objetivo de comparar o desempenho dos métodos conduzimos simulações apropriadas.

## Abstract

---

In this work we show the most important estimation methods for some item response models (both dichotomous and polichotomous). We discuss some proprieties of these methods. To compare the characteristic of these methods we conducted appropriate simulations.

---

# Capítulo 1

## Introdução

---

### 1.1 Introdução

O investimento dos países em educação constituiu uma das mais importantes vertentes na política de desenvolvimento. A pesquisa na área educacional permitiu que métodos estatísticos mais eficientes fossem criados o que possibilitou a obtenção de resultados, tais como a comparação de indivíduos de diferentes séries, os quais por sua vez, impulsionaram a implementação dos chamados Sistemas de Avaliação em Larga Escala, e.g. SAEB. Notadamente, os avanços da Teoria da Resposta ao Item (TRI), foram de suma importância em tais avaliações, por seu conjunto de características particulares que permitem obter resultados antes virtualmente impossíveis. Por exemplo, a comparação de resultados de grupos de alunos, submetidos a diferentes provas.

Dentro desse contexto, os processos de estimação nos vários modelos propostos na literatura da TRI constituem uma ferramenta imprescindível, uma vez que as aplicações dos mesmos dependem diretamente da eficiência desses métodos.

Nesse sentido, o presente trabalho propõe apresentar e discutir os mais importantes e utilizados métodos de estimação em algumas classes de modelos de resposta ao item.

### 1.2 Considerações sobre a Estimação Paramétrica

Basicamente, nos modelos da TRI, existem dois tipos de parâmetros que os caracterizam, a saber, **itens**, relacionados às questões e **habilidades** (ou proficiências), relacionados aos indivíduos. Quando se está frente a situação na qual se conhece os parâmetros dos itens e deseja-se estimar as habilidades de indivíduos (por exemplo, existem **bancos de itens** já calibrados (com os parâmetros conhecidos) e deseja-se classificar (selecionar) indivíduos, como por exemplo nos exames vestibulares, ou vice-versa, quando se conhece as habilidades dos indivíduos e deseja-se calibrar os itens (criação de **bancos de itens**), o problema de estimação torna-se relativamente simples (como será visto mais adiante), uma vez que o processo de estimação tornou-se de certo modo fáceis de serem implementados e também a escala na qual os parâmetros (itens ou habilidades) são estimados é determinada pelo conhecimento da escala na qual os parâmetros conhecidos foram estimados. Assim elimina-se o problema da indeterminação (não-identificabilidade) do

modelo [Andrade, Tavares & Valle (2000)]. O problema de não-identificabilidade reside no fato de que diferentes valores dos parâmetros podem levar a um mesmo valor da verossimilhança, o que, fatalmente, compromete o processo de obtenção das estimativas [Baker (1992)]. Além disso, os problemas relacionados aos algoritmos numéricos, tais como a não convergência dos processos iterativos, diminuem sensivelmente.

Por outro lado, quando desconhecemos ambos os tipos de parâmetros, surgem alguns problemas como, por exemplo, a necessidade de se determinar uma métrica (devido à não identificabilidade do modelo) e também a necessidade de se inverter matrizes de grandes dimensões, sendo que este último pode acarretar sérios problemas em termos de convergência computacional. Na verdade, o processo de estimação por **Máxima Verossimilhança Conjunta** (MVC) [Baker (1992)], é inaplicável para praticamente qualquer situação devido aos problemas mencionados anteriormente.

Na tentativa de contornar alguns desses problemas, Birnbaum (1968) propôs um processo *vai e volta* (“back-and-forth”) (uma reformulação no processo de MVC) o qual é iniciado com uma estimativa *grosseira* das habilidades (por exemplo o escore padronizado) e, subseqüentemente, os parâmetros dos itens são estimados considerando-se as habilidades conhecidas. Após esse passo, com esses valores, estima-se as habilidades considerando-se os parâmetros dos itens conhecidos [Birnbaum, faz uma série de imposições de modo a tornar a matriz Hessiana bloco diagonal, veja Baker (1992)]. Apesar de aparentemente o processo de estimação se tornar operacional, surgem alguns problemas. Primeiramente há a necessidade de em cada passo da estimação das habilidades assegurar que as mesmas estejam em alguma métrica. Birnbaum (1968) sugeriu que as mesmas fossem padronizadas, de modo a tê-las na escala (0;1) e também sugeriu algumas transformações das estimativas dos parâmetros dos itens, de modo a também tê-las nesta mesma escala [Baker (1992)]. Além disso, alguns autores encontraram diversos problemas nas propriedades assintóticas dos estimadores. Sabemos que, com os parâmetros dos itens conhecidos, os **Estimadores de Máxima Verossimilhança** (EMV) das habilidades convergem em probabilidade (propriedade de consistência) para os verdadeiros valores dos parâmetros [ver Sen & Singer (1993), por exemplo], e vice-versa, conhecida às habilidades, os EMV dos parâmetros dos itens convergem em probabilidade para os verdadeiros valores desses. Acontece que, nessa abordagem de estimação conjunta, os parâmetros dos itens são denominados **estruturais**, pois o número desses não aumenta com o aumento do número de indivíduos, o que não ocorre com as habilidades, denominadas de parâmetros **incidentais**. Esta nomenclatura é devida a Neyman & Scott (1948), que primeiro notaram este problema em um contexto diferente ao da TRI. Eles mostram que, quando os parâmetros estruturais (itens) são estimados juntamente com os parâmetros incidentais (habilidades), os EMV daqueles podem ser assintoticamente viesados, o que também foi notado por Wright (1977). E, mesmo sendo não-viciados, podem ser ineficientes (podem apresentar erros-padrão elevados). Além disso, na estimação dos parâmetros dos itens, as habilidades aparecem como parâmetros de perturbação [Harwell, Baker & Zwartz (1988)].

No entanto, quando o número de itens e o número de indivíduos crescem, os EMV de ambos os tipos de parâmetros (itens e habilidades) podem ser não viciados, como sugerido por Lord (1968). Alguns resultados numéricos obtidos por Lord (1975) e Swaminathan & Gifford (1983) reforçam a conjectura de que os EMV de ambos os tipos de parâmetros são não-viciados quando o número de itens e o número de indivíduos cresce, simultaneamente e numa mesma proporção. Contudo, o processo de estimação simultânea dos parâmetros pode se tornar difícil de ser operacionalizado considerando que o número de habilidades cresce com o tamanho da amostra [Harwell et al (1988)].

Muitos autores, incluindo Wright & Panchapakesan (1969), Wood, Wingerskyand & Lord (1976) e Kolakowski & Bock (1973a, 1973b), tentaram evitar o problema da falta de propriedades assintóticas desejáveis, através da suposição de que os indivíduos que tem o mesmo escore (número de respostas corretas) no teste ou o mesmo padrão de resposta (tipo de resposta), ou ainda, que tenham sido alocados provisoriamente a grupos homogêneos com relação as habilidades, têm de fato, a mesma habilidade. Sob esta ótica, o número de habilidades é finito e teoremas assintóticos padrão podem ser aplicados. Apesar desta abordagem (conhecida como solução de “efeitos fixos”) geralmente produzir estimativas razoáveis dos parâmetros dos itens, ela pode estar sujeita a vários vieses ou até mesmo tornar-se instável quando o número de itens é pequeno [Bock & Aitkin (1981)]. De qualquer forma, a suposição de que as habilidades são parâmetros fixos (no sentido de serem conhecidos) e finitos (com relação a quantidade dos mesmos), quando na verdade eles não são conhecidos (pois não conhecemos as verdadeiras habilidades dos indivíduos) e além disso, possuem uma distribuição de probabilidade associada a população da qual os indivíduos foram selecionados, é difícil de ser justificado do ponto de vista estatístico [Bock & Aitkin (1981)].

Com o intuito de contornar possíveis problemas de falta de propriedades assintóticas desejáveis dos EMV, Bock & Lieberman (1970), propuseram um método batizado de **Máxima Verossimilhança Marginal**. A idéia básica é considerar a existência de uma distribuição de probabilidade latente  $\Pi$ , associada às habilidades, e que os  $n$  indivíduos constituem uma amostra aleatória simples de uma população com essa distribuição [Andersen (1980)].

Em considerando essa idéia, ao invés de trabalharmos com a verossimilhança completa, trabalhamos, numa primeira etapa, com uma verossimilhança marginal (integrada), que possui as mesmas propriedades de uma verossimilhança genuína [Cordeiro(1992)] e também produz estimadores para os parâmetros dos itens que possuem as propriedades desejáveis dos estimadores de Máxima Verossimilhança [Baker(1992)].

Apesar desse método produzir EMV's com as propriedades assintóticas desejáveis, ele exige a estimação conjunta dos parâmetros de todos os itens, o que o torna inaplicável para testes com mais de 12 itens [Bock & Aitkin (1981)].

Em vista disso, Bock & Aitkin (1981) propuseram um aperfeiçoamento do método de MVM, o qual consiste em basicamente numa reformulação (uma forma mais adequada de se escrever)

das equações de estimação desenvolvidas por Bock & Lieberman, e o uso de uma adaptação do **algoritmo EM** [proposto por Dempster, Laird & Rubin (1977)], de tal modo que a matriz Hessiana, com relação aos parâmetros dos itens, pudesse se tornar, naturalmente, bloco-diagonal, permitindo assim que os parâmetros de cada item pudessem ser estimados em separado. Isso tornou o processo factível do ponto de vista computacional e ao mesmo tempo, preservou as propriedades assintóticas dos estimadores [Bock & Aitkin (1981)].

Baseado na proposta de Bock & Aitkin (1981), Mislevy (1986) desenvolveu um método bayesiano de estimação dos parâmetros dos itens denominado de **Estimação Bayesiana Marginal**. Basicamente este é uma extensão daquele, na medida que acopla à esperança da log-verossimilhança prioris convenientes. Dessa forma, a variante do **algoritmo EM** é aplicada na **distribuição a posteriori marginal**.

Para a estimação das habilidades, em geral, considera-se que as estimativas dos parâmetros dos itens são seus verdadeiros valores e, a partir daí, estima-se as habilidades a partir de uma **verossimilhança perfilada** [Baker(1992)].

Esses processos, em geral, são usados no ajuste dos diversos modelos tais como : **Policotômicos** [Van der Linden & Hambleton (1997)], **Multivariados** [Matos (2001)], **Multi-dimensionais** [Nojosa (2001)] e **Longitudinais** [Tavares (2001)].

### 1.3 Objetivos e Organização da Dissertação

O objetivo principal desta dissertação é o de apresentar e discutir com detalhes os principais métodos de estimação na TRI. Tentamos esclarecer questões que não estão muito claras na literatura. Para tanto escolhemos duas das principais classes de modelos : para respostas dicotômicas e policotômicas para uma única população. Esses modelos são bastante utilizados e possuem ampla importância na Área Educacional. Notadamente para os modelos policotômicos, a literatura disponível trata muito pouco a respeito dos processos de estimação. Serão abordadas as duas classes dos parâmetros que caracterizam estes modelos.

Os Capítulos 2 e 3 abordam a estimação clássica e bayesiana, respectivamente, nos Modelos Logísticos Unidimensionais de Respostas Dicotômicas. Os Capítulos 4 e 5, tratam da estimação clássica e bayesiana, respectivamente, no Modelo de Resposta Nominal. Os Capítulos 6 e 7, apresentam a estimação clássica e bayesiana no Modelo de Resposta Gradual, respectivamente. O Capítulo 8 apresenta sugestões recentes e métodos alternativos de estimação. No Capítulo 9 conduz-se algumas simulações pertinentes enquanto que o Capítulo 10 ressalta conclusões e sugestões para pesquisas futuras.



---

## Capítulo 2

# Modelos Logísticos Unidimensionais para Respostas Dicotômicas: Estimação por Máxima Verossimilhança

---

### 2.1 Introdução

Este capítulo dedica-se a apresentação e discussão dos métodos de estimação por Máxima Verossimilhança (MV) no modelo logístico unidimensional de 3 parâmetros para uma única população (MLUP3). Casos particulares desse modelo, são os modelos logísticos unidimensionais de 1 e 2 parâmetros, respectivamente MLUP1 e MLUP2. Basicamente serão abordados Métodos de Máxima Verossimilhança Convencional (utilizando-se as verossimilhanças genuínas) e Máxima Verossimilhança Marginal, mostrando-se os desenvolvimentos necessários e suas características, procurando elucidar algumas questões que não nos parecem estar muito claras na literatura pertinente.

Primeiramente, vamos introduzir o cenário no qual um conjunto de  $n$  indivíduos foi submetido a um teste (prova) de  $I$  itens (questões), dicotômicos ou dicotomizados (itens de múltipla escolha corrigidos como certo/errado). Considere também a seguinte variável aleatória de Bernoulli :

$$Y_{ij} = \begin{cases} 1, & \text{se o indivíduo } j \text{ responde corretamente ao item } i \\ 0, & \text{se o indivíduo } j \text{ responde incorretamente ao item } i, \end{cases}$$

com  $j = 1, \dots, n$  e  $i = 1, \dots, I$ .

O MLUP3 vem sendo utilizado com bastante freqüência nos últimos anos devido a grande quantidade de situações práticas que podem ser modeladas por ele, ou pelo menos, aproximadamente modeladas. O grande uso deste modelo está nos sistemas de avaliação educacional (SAEB por exemplo, [www.inep.gov.br](http://www.inep.gov.br)) devido aos importantes resultados obtidos através do mesmo. Um exemplo de aplicação deste modelo na análise do Vestibular 2000 da Universidade Federal do Ceará pode ser encontrado em Azevedo (2000 a). Sua formulação é dada através de (Função de Resposta ao Item (FRI)) :

$$P(Y_{ij} = 1|\theta_j, \zeta_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (2.1)$$

onde :

$a_i$  : parâmetro que representa a discriminação (inclinação) do item;

$b_i$  : parâmetro que representa a dificuldade (ponto de inflexão) do item;

$c_i$  : parâmetro que representa a chance de acerto ao item para indivíduos com a habilidade baixa, ou acerto casual;

$\theta_j$  : é a habilidade ou proficiência (traço latente) do indivíduo  $j$ .

$D$  : é um fator de escala igual a 1, se quisermos obter o resultado da função logística, ou 1,7 caso queiramos obter a aproximação para a distribuição normal .

Considere também as seguintes notações :

- $\mathbf{Y}_{.j} = (Y_{1j}, \dots, Y_{Ij})'$ , vetor aleatório de respostas do  $j$ -ésimo indivíduo aos  $I$  itens.
- $\mathbf{Y}_{..} = (\mathbf{Y}'_{.1}, \dots, \mathbf{Y}'_{.n})'$ , o conjunto integral de respostas.
- $\boldsymbol{\zeta}_i = (a_i, b_i, c_i)'$ , vetor dos parâmetros do  $i$ -ésimo item.
- $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1, \dots, \boldsymbol{\zeta}'_I)'$ , vetor dos parâmetros de todos os itens.
- $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)'$  vetor de habilidades de todos os indivíduos.

Vale ressaltar que existem problemas de indeterminação devido a não identificabilidade, relacionadas ao modelo (2.1), [para maiores detalhes ver Andrade, Tavares & Valle (2000)]. O modelo acima possui duas principais suposições que são imprescindíveis para o desenvolvimento dos processos de estimação :

**(S1)** as respostas oriundas de indivíduos diferentes são independentes.

**(S2)** os itens são respondidos de forma independente por cada indivíduo (independência local) dada a sua habilidade.

Não iremos nos deter em apresentar interpretações e características do modelo definido anteriormente, pois o nosso principal objetivo é o processo de estimação. Detalhes deste modelo podem ser encontrados em Andrade, Tavares & Valle (2000).

## 2.2 Estimação dos parâmetros dos itens

Nesta Seção apresentaremos o processo de estimação dos parâmetros dos itens. As expressões explicitadas aplicam-se a conjuntos de dados completos. Os dados omissos (falta de resposta de alguns indivíduos) ou respostas nulas (indivíduos que escolhem mais de um alternativa) serão considerados como respostas incorretas. Portanto, os procedimentos descritos podem ser aplicados.

### 2.2.1 Habilidades conhecidas

Por estar muito bem descrito na literatura disponível, simplesmente iremos apresentar as expressões finais pertinentes ao processo de estimação. Os desenvolvimentos inerentes podem ser encontrados em Baker (1992) e Andrade, Tavares & Valle (2000), por exemplo.

Pela independência entre as respostas dos diferentes indivíduos (S1) e a independência local (S2), podemos escrever a verossimilhança,  $L(\zeta) \equiv P(\mathbf{Y}_{..} = \mathbf{y}_{..} | \boldsymbol{\theta}, \zeta)$ , como

$$\begin{aligned} L(\zeta) &= \prod_{j=1}^n P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \theta_j, \zeta) \\ &= \prod_{j=1}^n \prod_{i=1}^I P(Y_{ij} = y_{ij} | \theta_j, \zeta_i), \end{aligned} \quad (2.2)$$

onde na última igualdade temos que a distribuição de  $Y_{ij}$  só depende de  $\zeta$  através de  $\zeta_i$  (pelo modelo). Usando a notação  $P_{ij} = P(Y_{ij} = 1 | \theta_j, \zeta_i)$  e  $Q_{ij} = 1 - P_{ij}$ , temos que

$$\begin{aligned} P(Y_{ij} = 1 | \theta_j, \zeta_i) &= P(Y_{ij} = 1 | \theta_j, \zeta_i)^{y_{ij}} P(Y_{ij} = 0 | \theta_j, \zeta_i)^{1-y_{ij}} \\ &= P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}}. \end{aligned}$$

Portanto, temos que a verossimilhança pode ser escrita como

$$L(\zeta) = \prod_{j=1}^n \prod_{i=1}^I P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}}. \quad (2.3)$$

Segue que a log-verossimilhança  $[(\ln L(\zeta) = l(\zeta))]$ , onde  $\ln$  denota o logaritmo natural pode ser escrita como

$$l(\zeta) = \sum_{j=1}^n \sum_{i=1}^I \{y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln Q_{ij}\},$$

O vetor escore (equações de estimação) resultante é dado por

$$\mathbf{S}(\zeta_i) = \frac{\partial l(\zeta)}{\partial \zeta_i} = \begin{bmatrix} D(1 - c_i) \sum_{j=1}^n (y_{ij} - P_{ij}) (\theta_j - b_i) W_{ij} \\ -Da_i(1 - c_i) \sum_{j=1}^n (y_{ij} - P_{ij}) W_{ij} \\ \sum_{j=1}^n \left\{ (y_{ij} - P_{ij}) \frac{W_{ij}}{P_{ij}^*} \right\} \end{bmatrix}, \quad (2.4)$$

com  $W_{ij} = \frac{P_{ij}^* Q_{ij}^*}{P_{ij} Q_{ij}}$ ,  $P_{ij}^* = \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1}$  e  $Q_{ij}^* = 1 - P_{ij}^*$ .

Como o sistema de equações descrito em (2.4) não possui solução explícita, devemos utilizar algum método iterativo. Os dois métodos iterativos mais usados em problemas de estimação paramétrica e que também são muito usados na TRI, são o Método de Newton-Raphson [ver Isaac & Keller(1996), por exemplo] e Escore de Fisher [ver Rao(1973), por exemplo]. Ambos necessitam do cálculo da **matriz Hessiana (Informação de Fisher)**. Estas são dadas respectivamente por

$$\begin{aligned} \mathbf{H}(\zeta_i) &= \sum_{j=1}^n \left\{ \left( \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right) (P_{ij}^* Q_{ij}^*) \mathbf{H}_{ij} - \left( \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right)^2 (P_{ij}^* Q_{ij}^*)^2 \mathbf{h}_{ij} \mathbf{h}'_{ij} \right\} \\ &= \sum_{j=1}^n (y_{ij} - P_{ij}) W_{ij} \left\{ \mathbf{H}_{ij} - W_{ij} (y_{ij} - P_{ij}) \mathbf{h}_{ij} \mathbf{h}'_{ij} \right\} \end{aligned} \quad (2.5)$$

e

$$\begin{aligned} \mathbf{I}(\zeta_i) &= \mathbb{E} \{ -\mathbf{H}(\zeta_i) \} \\ &= - \sum_{j=1}^n \left\{ W_{ij} \mathbb{E} (Y_{ij} - P_{ij}) \mathbf{H}_{ij} - W_{ij}^2 \mathbb{E} (Y_{ij} - P_{ij})^2 \mathbf{h}_{ij} \mathbf{h}'_{ij} \right\} \\ &= \sum_{j=1}^n P_{ij} Q_{ij} \frac{(P_{ij}^* Q_{ij}^*)^2}{(P_{ij} Q_{ij})^2} \mathbf{h}_{ij} \mathbf{h}'_{ij} = \sum_{j=1}^n P_{ij}^* Q_{ij}^* W_{ij} \mathbf{h}_{ij} \mathbf{h}'_{ij}, \end{aligned} \quad (2.6)$$

em que

$$\mathbf{h}_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left( \frac{\partial P_{ij}}{\partial \zeta_i} \right) = \begin{pmatrix} D(1 - c_i)(\theta_j - b_i) \\ -Da_i(1 - c_i) \\ \frac{1}{P_{ij}^*} \end{pmatrix}$$

e

$$\begin{aligned} \mathbf{H}_{ij} &= (P_{ij}^* Q_{ij}^*)^{-1} \left( \frac{\partial^2 P_{ij}}{\partial \zeta_i \partial \zeta_i'} \right) \\ &= \begin{pmatrix} D^2(1 - c_i)(\theta_j - b_i)^2(1 - 2P_{ij}^*) & \cdot & \cdot \\ -D(1 - c_i) \left\{ 1 + Da_i(\theta_j - b_i)(1 - 2P_{ij}^*) \right\} & D^2 a_i^2 (1 - c_i)(1 - 2P_{ij}^*) & \cdot \\ -D(\theta_j - b_i) & Da_i & 0 \end{pmatrix}. \end{aligned}$$

Dessa forma, considerando  $\hat{\zeta}_i^{(t)}$  uma estimativa de  $\zeta_i$  na iteração  $t$ , podemos definir os procedimentos de Newton-Raphson/Escore de Fisher como :

**Newton-Raphson**

$$\hat{\zeta}_i^{(t+1)} = \hat{\zeta}_i^{(t)} - [\mathbf{H}(\hat{\zeta}_i^{(t)})]^{-1} \mathbf{S}(\hat{\zeta}_i^{(t)}) \quad (2.7)$$

**Escore de Fisher**

$$\hat{\zeta}_i^{(t+1)} = \hat{\zeta}_i^{(t)} + [\mathbf{I}(\hat{\zeta}_i^{(t)})]^{-1} \mathbf{S}(\hat{\zeta}_i^{(t)}) \quad (2.8)$$

$t = 0, 1, 2, \dots$

Na literatura existem algumas sugestões para os valores iniciais dos processos iterativos. Em geral estes são baseados em relações entre os parâmetros da TRI e da Teoria Clássica de Medidas [Lord & Novick (1968)]. Detalhes desses valores iniciais podem ser encontrados em Andrade, Tavares & Valle (2000).

**Agrupamento das habilidades**

Um procedimento alternativo para a estimação dos parâmetros dos itens, quando as habilidades dos indivíduos são conhecidas, é considerar um agrupamento dessas em categorias. Isso pode acarretar em algumas vantagens computacionais, haja vista que o número de *laços* nos algoritmos pode diminuir sensivelmente, sobretudo quando estamos lidando com um grande número de indivíduos. Dito isto, podemos considerar o agrupamento das habilidades, por exemplo, definindo um conjunto de  $q$  intervalos cujos valores médios (ou outra medida central), sejam denotados por  $\bar{\theta}_l$ ,  $l = 1, \dots, q$ . Podemos considerar, para fins de desenvolvimento, que todos os indivíduos pertencentes à categoria  $l$ , têm habilidade igual (ou muito próxima) a  $\bar{\theta}_l$ , isso, como já foi dito, pode reduzir bastante o esforço computacional. Além disso, o desenvolvimento com esta abordagem, será útil mais adiante quando estivermos na situação em que ambos os tipos de parâmetros são desconhecidos.

De uma forma geral, consideramos que  $q$  grupos de  $f_{il}$  indivíduos com habilidades conhecidas (em torno de)  $\bar{\theta}_l$  são selecionados ao acaso, segundo um processo de amostragem aleatória simples sem reposição de uma população  $\Pi$  em estudo para responder a um determinado teste. Sejam  $f_{il}$  e  $r_{il}$  o número de indivíduos com nível de habilidade em torno de  $\bar{\theta}_l$  que respondem ao item  $i$  e, dentre esses indivíduos, o número daqueles que respondem corretamente a esse mesmo item, respectivamente. Nota-se que podemos ignorar o índice  $i$  da quantidade  $f_{il}$ , pois além de podermos estimar os parâmetros de cada item em separado, a amostra de indivíduos será submetida a todos os itens, pelo menos em princípio. Entretanto, na prática, é comum que alguns indivíduos não respondam (ou anulem de outra forma) alguns itens, ou ainda pela própria natureza do sistema de avaliação, nem todos os itens são apresentados a cada um dos indivíduos, como por exemplo, nos cadernos de provas construídos a partir do **planejamento em blocos incompletos balanceados** (BIB)[ver Azevedo (2000 b)].

Dessa forma consideraremos essa abordagem mais ampla. Pela independência entre as respostas dos indivíduos, temos que os  $r_{il}$ 's são variáveis aleatórias independentes, cada qual seguindo uma distribuição **Binomial** de parâmetros  $f_{il}$  e  $P_{il}$  sendo este último como definido em (2.1), com  $\theta_j$ , sendo substituído por  $\bar{\theta}_l$ . Assim, temos que a verossimilhança pode ser escrita como

$$L(\zeta) = \prod_{i=1}^I \prod_{l=1}^q \left\{ \binom{f_{il}}{r_{il}} P_{il}^{r_{il}} Q_{il}^{f_{il}-r_{il}} \right\}. \quad (2.9)$$

Por sua vez, a log-verossimilhança é dada por

$$l(\zeta) = \sum_{l=1}^q \sum_{i=1}^I \ln \binom{f_{il}}{r_{il}} + \sum_{l=1}^q \sum_{i=1}^I \{r_{il} \ln P_{il} + (f_{il} - r_{il}) \ln Q_{il}\}. \quad (2.10)$$

Seguindo desenvolvimento análogo ao caso de não agrupamento das habilidades, temos que o vetor escore, para os parâmetros de um determinado item, é dado por :

$$\begin{aligned} S(\zeta_i) &= \frac{\partial L(\zeta)}{\partial \zeta_i} \\ &= \begin{bmatrix} S(a_i) \\ S(b_i) \\ S(c_i) \end{bmatrix} = \begin{bmatrix} D \sum_{l=1}^q \{r_{il} - f_{il}\} (1 - c_i) (\bar{\theta}_l - b_i) W_{il} \\ -D a_i (1 - c_i) \sum_{l=1}^q \{r_{il} - f_{il} P_{il}\} W_{il} \\ \sum_{l=1}^q \{r_{il} - f_{il} P_{il}\} \frac{W_{il}}{P_{il}^*} \end{bmatrix}. \end{aligned} \quad (2.11)$$

De maneira análoga, temos que a Matriz Hessiana e a Informação de Fisher são dadas por :

**Matriz Hessiana**

$$H(\zeta_i) = \sum_{l=1}^q \left[ W_{il} \left\{ (r_{il} - f_{il} P_{il}) \mathbf{H}_{il} - W_{il} (r_{il} - 2r_{il} P_{il} + f_{il} P_{il}^2) \mathbf{h}_{il} \mathbf{h}'_{il} \right\} \right], \quad (2.12)$$

**Informação de Fisher**

$$I(\zeta_i) = \sum_{l=1}^q f_{il} P_{il}^* Q_{il}^* \mathbf{h}_{il} \mathbf{h}'_{il}, \quad (2.13)$$

com

$$\mathbf{h}_{il} = (P_{il}^* Q_{il}^*)^{-1} \left( \frac{\partial P_{il}}{\partial \zeta_i} \right) = \begin{pmatrix} D(1 - c_i)(\bar{\theta}_l - b_i) \\ -Da_i(1 - c_i) \\ \frac{1}{P_{il}^*} \end{pmatrix}$$

e

$$\begin{aligned} \mathbf{H}_{il} &= (P_{il}^* Q_{il}^*)^{-1} \left( \frac{\partial^2 P_{il}}{\partial \zeta_i \partial \zeta_i'} \right) \\ &= \begin{pmatrix} D^2(1 - c_i)(\bar{\theta}_l - b_i)^2(1 - 2P_{il}^*) & \cdot & \cdot \\ -D(1 - c_i)\{1 + a_i(\bar{\theta}_l - b_i)(1 - 2P_{il}^*)\} & D^2 a_i^2(1 - c_i)(1 - 2P_{il}^*) & \cdot \\ -D(\bar{\theta}_l - b_i) & Da_i & 0 \end{pmatrix}. \end{aligned}$$

Descreveremos na próxima seção o procedimento de estimação por Máxima Verossimilhança Marginal.

### 2.2.2 Habilidades desconhecidas

O processo de estimação clássico dos parâmetros dos itens (quando desconhecemos as habilidades), mais usado e com propriedades importantes, além de ser factível computacionalmente, é o método de **Máxima Verossimilhança Marginal - MVM** via abordagem de Bock & Aitkin [Bock & Aitkin (1981)], que é um aperfeiçoamento da abordagem de Bock & Lieberman [Bock & Lieberman (1970)]. O desenvolvimento usado nesta parte do trabalho, baseia-se em Baker(1992) e Andrade, Tavares & Valle(2000), sendo que neste primeiro pode-se encontrar alguns detalhes deste processo de estimação, e no segundo têm-se uma ótima referência em língua portuguesa.

A idéia básica, como já foi dito, é considerar a existência de uma distribuição de probabilidade latente  $\Pi$ , associada as habilidades, e que os  $n$  indivíduos constituem uma amostra aleatória simples de uma população com essa distribuição [Andersen(1980)]. É importante ressaltar que não se está aplicando nenhum argumento bayesiano. A densidade  $g(\theta)$  pode realmente ser considerada no sentido de se realizar um experimento de retirar um indivíduo dessa população e observar seu traço latente  $\theta$ .

Em geral considera-se uma densidade  $g(\theta|\boldsymbol{\eta})$ , duplamente diferenciável, com as componentes de  $\boldsymbol{\eta}$  conhecidas e finitas [vetor de parâmetros associados a  $g(\cdot)$ ]. Para o caso mais explorado, em que  $\theta$  tem distribuição normal, temos que  $\boldsymbol{\eta} = (\mu, \sigma^2)$ , onde  $\mu$  é a média e  $\sigma^2$  é a variância das habilidades dos indivíduos pertencentes a  $\Pi$ . Vale ressaltar que a distribuição normal não é a única opção.

Com as notações definidas anteriormente, temos que a probabilidade marginal de um indivíduo  $j$  apresentar um determinado padrão de respostas  $\mathbf{y}_j$  é dada por

$$\begin{aligned} P(Y_{.j} = \mathbf{y}_{.j} | \zeta, \boldsymbol{\eta}) &\equiv P(\mathbf{Y}_{.j} | \zeta, \boldsymbol{\eta}) = \int_{\mathbb{R}} P(Y_{.j} = \mathbf{y}_{.j} | \theta, \zeta, \boldsymbol{\eta}) g(\theta | \boldsymbol{\eta}) d\theta \\ &= \int_{\mathbb{R}} P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \boldsymbol{\eta}) d\theta, \end{aligned}$$

onde na última igualdade temos que a distribuição de  $\mathbf{Y}_{.j}$  não é função de  $\boldsymbol{\eta}$ .

Pela independência entre as respostas de diferentes indivíduos, temos que a verossimilhança é dada por

$$L(\zeta, \boldsymbol{\eta}) = P(\mathbf{Y}_{..} | \zeta, \boldsymbol{\eta}) = \prod_{j=1}^n P(\mathbf{Y}_{.j} | \zeta, \boldsymbol{\eta}),$$

e conseqüentemente, a logverossimilhança dada por

$$l(\zeta, \boldsymbol{\eta}) = \sum_{j=1}^n \ln P(\mathbf{Y}_{.j} | \zeta, \boldsymbol{\eta}). \quad (2.14)$$

Muito embora iremos manter a componente  $\boldsymbol{\eta}$  na log-verossimilhança definida em (2.14), ressaltamos que neste procedimento, fixamos essa componente a fim de resolver o problema da indeterminação.

Com os desenvolvimentos descritos em Baker (1992) e Andrade, Tavares & Valle (2000), chegamos as seguintes equações de estimação :

$$a_i : (1 - c_i) \sum_{j=1}^n \int_{\mathbb{R}} [(y_{ij} - P_i) (\theta - b_i) W_i] g_j^*(\theta) = 0 \quad (2.15)$$

$$b_i : -a_i (1 - c_i) \sum_{j=1}^n \int_{\mathbb{R}} [(y_{ij} - P_i) W_i] g_j^*(\theta) = 0 \quad (2.16)$$

$$c_i : \sum_{j=1}^n \int_{\mathbb{R}} \left[ (y_{ij} - P_i) \frac{W_i}{P_i^*} \right] g_j^*(\theta) = 0, \quad (2.17)$$

onde  $W_i = \frac{P_i^* Q_i^*}{P_i Q_i}$ ,  $P_i = c_i + (1 - c_i) \{1 + e^{-Da_i(\theta - b_i)}\}^{-1}$ ,  $P_i^* = \{1 + e^{-Da_i(\theta - b_i)}\}^{-1}$ ,  $Q_i = 1 - P_i$ ,  $Q_i^* = 1 - P_i^*$  e  $g_j^*(\theta) = \frac{P(\mathbf{Y}_{.j} | \zeta, \theta) g(\theta | \boldsymbol{\eta})}{\int P(\mathbf{Y}_{.j} | \zeta, \theta) g(\theta | \boldsymbol{\eta}) d\theta}$ .

As quais não possuem solução explícita.

### Métodos de Integração Numérica - Distribuição de pontos de quadratura

Podemos notar que as integrais das equações de estimação descritas em (2.15) a (2.17) não possuem solução explícita, sendo necessário o emprego de métodos (numéricos) de aproximação de integrais. Se  $g(\theta | \boldsymbol{\eta})$  é uma função contínua (densidade) com integral finita (momentos finitos)



ela pode ser aproximada com qualquer grau de precisão desejado por uma distribuição discreta com um número finito de pontos (i.e., por um histograma). O mais usado na TRI é o chamado método de Hermite-Gauss (ou quadratura gaussiana), que consiste basicamente em substituir o cálculo da integral, por uma soma finita das áreas de, digamos,  $q$  retângulos [ver Hilderbrand (1956) e Baker(1992)]. Os pontos médios desses retângulos, digamos  $\bar{\theta}_l$ ,  $l = 1, \dots, q$ , são chamados de nós (ou pontos de quadratura). Cada um desses pontos de quadratura tem um peso  $A_l \equiv A(\bar{\theta}_l)$  associado que é composto pelo valor da função (altura)  $g(\bar{\theta}_l|\boldsymbol{\eta})$  e o respectivo comprimento desse intervalo  $\Delta_l$ , de tal modo que  $A_l = g(\bar{\theta}_l|\boldsymbol{\eta}) \times \Delta_l$ . Para obtenção desses valores ver Hilderbrand (1956), pp. 327-330. Convém dizer que não é necessário que  $g(\bar{\theta}_l|\boldsymbol{\eta})$  tenha distribuição normal, nem necessariamente contínua, em geral, ela pode ser definida empiricamente [Mislevy & Bock (1982)]. Dessa forma, substituindo as aproximações por pontos quadratura nas equações (2.15) a (2.17), temos que :

$$a_i : (1 - c_i) \sum_{j=1}^n \sum_{l=1}^q [(y_{ij} - P_{il}) (\bar{\theta}_l - b_i) W_{il}] g_j^*(\bar{\theta}_l) = 0 \quad (2.18)$$

$$b_i : -a_i (1 - c_i) \sum_{j=1}^n \sum_{l=1}^q [(y_{ij} - P_{il}) W_{il}] g_j^*(\bar{\theta}_l) = 0 \quad (2.19)$$

$$c_i : \sum_{j=1}^n \sum_{l=1}^q \left[ (y_{ij} - P_{il}) \frac{W_{il}}{P_{il}^*} \right] g_j^*(\bar{\theta}_l) = 0 \quad (2.20)$$

com  $P_{il}$  sendo definido como  $P_{ij}$  substituindo  $\theta_j$  por  $\bar{\theta}_l$ ,  $Q_{il} = 1 - P_{il}$  e  $W_{il} = \frac{P_{il}^* Q_{il}^*}{P_{il} Q_{il}}$ ,

onde

$$g_j^*(\bar{\theta}_l) \simeq \frac{P(\mathbf{Y}_{.j}|\bar{\theta}_l, \boldsymbol{\zeta}) A_l}{\sum_{l=1}^q P(\mathbf{Y}_{.j}|\bar{\theta}_l, \boldsymbol{\zeta}) A_l} \quad (2.21)$$

e

$$P(\mathbf{Y}_{.j}|\bar{\theta}_l, \boldsymbol{\zeta}) = \prod_{i=1}^I \left[ P_{il}^{y_{ij}} Q_{il}^{1-y_{ij}} \right],$$

$$P(\mathbf{Y}_{ij}|\boldsymbol{\zeta}, \boldsymbol{\eta}) \simeq \sum_{l=1}^q P(\mathbf{Y}_{.j}|\bar{\theta}_l, \boldsymbol{\zeta}) g(\bar{\theta}_l|\boldsymbol{\eta}) \Delta_l \simeq \sum_{l=1}^q P(\mathbf{Y}_{.j}|\bar{\theta}_l, \boldsymbol{\zeta}) g(\bar{\theta}_l|\boldsymbol{\eta}),$$

desde que  $\Delta_l \cong 1$ .

Dessa forma, se continuássemos trabalhando com as equações (2.18) a (2.20) e desenvolvêssemos algum dos dois processos iterativos, estaríamos utilizando a abordagem de Bock & Lieberman, a qual conduz a necessidade de estimação conjunta dos parâmetros de todos os itens, e, muito

embora, produza EMV com propriedades assintóticas desejáveis, é inaplicável para testes com mais de 12 itens [Bock & Aitkin(1981)]. Apresentaremos a seguir a abordagem de Bock & Aitkin.

### Abordagem de Bock & Aitkin

Basicamente Bock & Aitkin propuseram uma reformulação conveniente nas equações de estimação dos parâmetros dos itens e o uso de uma adaptação do algoritmo EM [proposto por Dempster, Laird & Rubin (1977)], de tal modo que as derivadas cruzadas da logverossimilhança com relação aos parâmetros de diferentes itens se tornassem naturalmente nulas, permitindo assim que os parâmetros de cada item possam ser estimados em separado, tornando o processo factível do ponto de vista computacional e ao mesmo tempo, preservando as propriedades assintótica dos estimadores [Bock & Aitkin (1981)].

A reformulação proposta nas equações, que nada mais é que uma outra forma de escrevê-las, está descrita a seguir. Partindo de (2.18) a (2.20), temos que :

**Discriminação :**  $a_i$

$$\begin{aligned} a_i & : (1 - c_i) \sum_{l=1}^q \left[ \left( \sum_{j=1}^n y_{ij} g_j^* (\bar{\theta}_l) - P_{il} \sum_{j=1}^n g_j^* (\bar{\theta}_l) \right) (\bar{\theta}_l - b_i) W_{il} \right] \\ & = (1 - c_i) \sum_{l=1}^q [(\bar{r}_{il} - \bar{f}_{il} P_{il}) (\bar{\theta}_l - b_i) W_{il}]. \end{aligned}$$

**Dificuldade :**  $b_i$

$$\begin{aligned} b_i & : -a_i (1 - c_i) \sum_{l=1}^q \left[ \left( \sum_{j=1}^n y_{ij} g_j^* (\bar{\theta}_l) - P_{il} \sum_{j=1}^n g_j^* (\bar{\theta}_l) \right) W_{il} \right] \\ & = -a_i (1 - c_i) \sum_{l=1}^q [(\bar{r}_{il} - \bar{f}_{il} P_{il}) W_{il}]. \end{aligned}$$

**Acerto casual :**  $c_i$

$$\begin{aligned} c_i & : \sum_{l=1}^q \left[ \left( \sum_{j=1}^n y_{ij} g_j^* (\bar{\theta}_l) - P_{il} \sum_{j=1}^n g_j^* (\bar{\theta}_l) \right) \frac{W_{il}}{P_{il}^*} \right] \\ & = \sum_{l=1}^q \left[ (\bar{r}_{il} - \bar{f}_{il} P_{il}) \frac{W_{il}}{P_{il}^*} \right]. \end{aligned}$$

Resumidamente, temos que as equações de estimação são dadas por :

$$a_i : (1 - c_i) \sum_{l=q}^n [(\bar{r}_{il} - \bar{f}_{il}P_{il}) (\bar{\theta}_l - b_i) W_{il}] = 0 \quad (2.22)$$

$$b_i : -a_i (1 - c_i) \sum_{l=1}^q [(\bar{r}_{il} - \bar{f}_{il}P_{il}) W_{il}] = 0 \quad (2.23)$$

$$c_i : \sum_{l=1}^q \left[ (\bar{r}_{il} - \bar{f}_{il}P_{il}) \frac{W_{il}}{P_{il}^*} \right] = 0, \quad (2.24)$$

onde :

$$\bar{r}_{il} = \sum_{j=1}^n y_{ij} g_j^*(\bar{\theta}_l) \quad , \quad \bar{f}_{il} = \sum_{j=1}^n g_j^*(\bar{\theta}_l). \quad (2.25)$$

Aqui neste ponto é importante interpretar detalhadamente as expressões descritas em (2.25), pois elas representam quantidades importantes que são pontos cruciais para o desenvolvimento e entendimento do uso da adaptação do algoritmo EM, que será descrito mais adiante.

Primeiramente notemos que a expressão (2.21) representa a distribuição a posteriori das habilidades (para o  $j$ -ésimo indivíduo condicionado ao seu padrão de resposta, parâmetros dos itens e populacionais). Basicamente o que a distribuição a posteriori faz é combinar a informação da verossimilhança com a priori e distribuir os indivíduos, condicionado aos vetores de respostas aos itens, ao longo da escala de habilidade na proporção da probabilidade a posteriori dos indivíduos pertencerem a algum ponto da escala de habilidade. Observando novamente a expressão (2.21) notamos que a mesma está representando, de forma empírica (discretizada), a probabilidade a posteriori do  $j$ -ésimo indivíduo apresentar habilidade em torno (do nível)  $\bar{\theta}_l$ , ou seja, condicionado a seu respectivo vetor de respostas, os parâmetros dos itens e os parâmetros relativos às priors das habilidades.

Consideremos, agora, a seguinte variável de Bernoulli :

$$X_{(i)jl} = \begin{cases} 1, & \text{se o indivíduo } j \text{ responde (é submetido) ao} \\ & \text{item } i \text{ e possui habilidade em torno de } \bar{\theta}_l \\ 0, & \text{caso contrário.} \end{cases}$$

Sejam  $f_{il}$  e  $r_{il}$  como descritos anteriormente, ou seja, o número de indivíduos submetidos (que respondem) ao item  $i$ , e desses indivíduos, o número daqueles que responderam corretamente ao item  $i$ , respectivamente. Assim, podemos reescrevê-las do seguinte modo:

$$f_{il} = \sum_{j=1}^n X_{(i)jl}; \quad r_{il} = \sum_{j=1}^n Y_{ij} X_{(i)jl}. \quad (2.26)$$

Por outro lado, condicionado as parâmetros dos itens (estimativas), as respostas dos indivíduos e aos parâmetros da distribuição a priori das habilidades temos que

$$X_{(i)jl} | (\mathbf{Y}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) \sim \text{Bernoulli} (g_j^* (\bar{\theta}_l)) ; P (X_{(i)jl} = 1) \equiv g_j^* (\bar{\theta}_l) .$$

Dessa forma, as esperanças a posteriori das quantidades definidas em (2.26), são respectivamente :

$$\mathbb{E} (F_{il} | \mathbf{Y}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{j=1}^n \mathbb{E} (X_{(i)jl} | \mathbf{Y}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{j=1}^n g_j^* (\bar{\theta}_l) = \bar{f}_{il} \quad (2.27)$$

e

$$\mathbb{E} (R_{il} | \mathbf{Y}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{j=1}^n \mathbb{E} (y_{ij} X_{(i)jl} | \mathbf{Y}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{j=1}^n y_{ij} g_j^* (\bar{\theta}_l) = \bar{r}_{il}. \quad (2.28)$$

Dessa forma, notamos que  $\bar{f}_{il}$  e  $\bar{r}_{il}$  representam, respectivamente, o número esperado de indivíduos que respondem ao item  $i$  com habilidade em torno de  $\bar{\theta}_l$ , e dentre esses mesmos indivíduos, o número esperado daqueles que respondem corretamente ao supracitado item. Como veremos adiante, e foi dito anteriormente, essas duas quantidades desempenham papéis fundamentais na construção da variante do algoritmo EM.

De posse desses resultados estamos aptos para descrevermos a adaptação do algoritmo EM em nosso problema.

### Aplicação do algoritmo EM

De um modo geral o algoritmo EM é um processo iterativo para a determinação de estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade na presença de variáveis aleatórias não observadas, podendo ser estendido para o caso de estimação Bayesiana (moda a posteriori) [Dempster, Laird & Rubin (1977)]. No caso da TRI, o objetivo é obter estimativas de  $\boldsymbol{\zeta}$  na presença das variáveis não observadas  $\boldsymbol{\theta}$ . Neste caso,  $\mathbf{Y}_{..}$  representa o vetor de dados incompletos e  $(\mathbf{Y}_{..}, \boldsymbol{\theta})$  o vetor de dados completos. Seja  $f(\mathbf{Y}_{..}, \boldsymbol{\theta} | \boldsymbol{\zeta})$  a densidade conjunta do dados completos. Se  $\hat{\boldsymbol{\zeta}}^{(t)}$  é uma estimativa de  $\boldsymbol{\zeta}$  na iteração  $t$ , então os passos EM para obtenção de  $\hat{\boldsymbol{\zeta}}^{(t+1)}$  são

**Passo E:** Calcular  $E[\ln f(\mathbf{Y}_{..}, \boldsymbol{\theta} | \boldsymbol{\zeta}) | \mathbf{Y}_{..}, \hat{\boldsymbol{\zeta}}^{(t)}]$ .

**Passo M:** Obter  $\hat{\boldsymbol{\zeta}}^{(t+1)}$  que maximiza a função do Passo E.

No Passo M a maximização pode ser feita via algoritmo Newton-Raphson ou Escore de Fisher, a qual é feita em separado para cada vetor de parâmetros de cada item, como será visto mais adiante.

Basicamente, há três formas do algoritmo EM, distinguidas pela relação entre a função (densidade) de probabilidade (MRI) e a forma da família exponencial, [ver Dempster, Laird & Rubin (1977)]. No caso do MLUP3 a terceira forma do algoritmo EM deve ser aplicada pelo fato de que o mesmo não pertence à família exponencial, muito embora o MLUP1 pertença. Na verdade, notemos que  $f(\mathbf{Y}_{..}, \boldsymbol{\theta}|\boldsymbol{\zeta})$  em geral não possui forma conhecida pois

$$\begin{aligned} f(\mathbf{Y}_{..}, \boldsymbol{\theta}|\boldsymbol{\zeta}) &= f(\mathbf{Y}_{..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\zeta}, \boldsymbol{\eta}) \\ &= L(\boldsymbol{\zeta}, \boldsymbol{\theta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) = \prod_{i=1}^I \prod_{j=1}^n \left\{ P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}} \right\} \left\{ \prod_{j=1}^n g(\theta_j|\boldsymbol{\eta}) \right\}, \end{aligned}$$

de onde temos que  $L(\boldsymbol{\zeta}, \boldsymbol{\theta})$  (a verossimilhança) é conhecida, mas  $g(\theta_j|\boldsymbol{\eta}) \equiv g(\boldsymbol{\theta}|\boldsymbol{\eta})$  (a priori) embora seja atribuída, na prática, não a conhecemos.

Para descrever brevemente o algoritmo EM aplicado à TRI, comecemos supondo que as habilidades estão restritas a um conjunto de  $q$  valores,  $\bar{\theta}_l$ ,  $l = 1, \dots, q$ , com proporções  $\pi_l$ ,  $l = 1, \dots, q$ . Essa suposição pode ser feita porque as aproximações de integrais são realizadas por métodos de quadratura, e os valores  $\bar{\theta}_l$  corresponderão aos pontos de quadratura. Seja  $f_{il}$  como já descrito anteriormente,  $\mathbf{f}_i = (f_{i1}, \dots, f_{iq})'$ , com  $\sum_{l=1}^q f_{il} = n_{(i)}$ ,  $\mathbf{f} = (\mathbf{f}'_1, \dots, \mathbf{f}'_I)'$ . Similarmente, seja  $r_{il}$  como também já descrito,  $\mathbf{r}_i = (r_{i1}, \dots, r_{iq})'$  e  $\mathbf{r} = (\mathbf{r}'_1, \dots, \mathbf{r}'_I)'$ . Notamos que as quantidades  $f_{il}$  e  $r_{il}$  não são conhecidas. Essa é a grande vantagem do algoritmo EM, onde  $f_{il}$  e  $r_{il}$  podem ser tratadas como quantidades não observadas ( conhecidas como **dados artificiais**) [Baker (1992)].

Se os  $n_{(i)}$  indivíduos que responderão ao item  $i$  são selecionados ao acaso da população, a probabilidade conjunta que os  $f_{il}$  indivíduos tenham habilidades  $\bar{\theta}_l$ ,  $l = 1, \dots, q$ , é dada pela **distribuição multinomial**:

$$P(\mathbf{F}_i = \mathbf{f}_i|\boldsymbol{\pi}) \equiv P(\mathbf{f}_i|\boldsymbol{\pi}) = \frac{n_{(i)}!}{\prod_{l=1}^q f_{il}!} \prod_{l=1}^q \pi_l^{f_{il}}, \quad i = 1, \dots, I.$$

Dados  $f_{il}$  e  $\bar{\theta}_l$ , a probabilidade de ocorrerem  $r_{il}$  acertos ao item  $i$  dentre as  $f_{il}$  tentativas (respostas) dos indivíduos com habilidade  $\bar{\theta}_l$  é

$$P(R_{il} = r_{il}|f_{il}, \bar{\theta}_l) \equiv P(r_{il}|f_{il}, \bar{\theta}_l) = \binom{f_{il}}{r_{il}} P_{il}^{r_{il}} Q_{il}^{f_{il}-r_{il}}.$$

A probabilidade conjunta de  $\mathbf{f}$  e  $\mathbf{r}$ , dados  $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1, \dots, \bar{\theta}_q)'$  e  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_q)'$ , é

$$\begin{aligned}
P(\mathbf{F} = \mathbf{f}, \mathbf{R} = \mathbf{r} | \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) &\equiv P(\mathbf{f}, \mathbf{r} | \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) = P(\mathbf{r} | \mathbf{f}, \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) P(\mathbf{f} | \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) \\
&= P(\mathbf{r} | \mathbf{f}, \bar{\boldsymbol{\theta}}) P(\mathbf{f} | \boldsymbol{\pi}) \\
&= \left\{ \prod_{i=1}^I \prod_{l=1}^q P(r_{il} | f_{il}, \bar{\theta}_l) \right\} \left\{ \prod_{i=1}^I P(\mathbf{f}_i | \boldsymbol{\pi}) \right\}.
\end{aligned}$$

Pelo critério da fatoração, temos que  $(\mathbf{f}', \mathbf{r}')$  é uma estatística suficiente para os dados completos  $(\mathbf{Y}'_{..}, \boldsymbol{\theta}')$  [Baker (1992)]. Seque que a log-verossimilhança para os dados completos é :

$$\begin{aligned}
\ln L(\boldsymbol{\zeta}) &= \ln P(\mathbf{f} | \boldsymbol{\pi}) + \sum_{i=1}^I \sum_{l=1}^q \ln P(r_{il} | f_{il}, \bar{\theta}_l) \\
&= \ln P(\mathbf{f} | \boldsymbol{\pi}) + \sum_{i=1}^I \sum_{l=1}^q \left\{ \ln \binom{f_{il}}{r_{il}} + r_{il} \ln P_{il} + (f_{il} - r_{il}) \ln Q_{il} \right\} \\
&= C + \sum_{l=1}^q \sum_{i=1}^I \{ r_{il} \ln P_{il} + (f_{il} - r_{il}) \ln Q_{il} \},
\end{aligned}$$

onde  $C = \ln P(\mathbf{f} | \boldsymbol{\pi}) + \sum_{i=1}^I \sum_{l=1}^q \ln \binom{f_{il}}{r_{il}}$  é constante com relação a  $\boldsymbol{\zeta}$ . Temos que  $(\mathbf{f}', \mathbf{r}')$  são não-observáveis, mas tomando a esperança da log-verossimilhança, condicionada em  $\mathbf{Y}_{..}$  e  $\boldsymbol{\zeta}$ , usando (2.27) e (2.28) e considerando a seguinte notação

$$\bar{C} = E[C | \mathbf{Y}_{..}, \boldsymbol{\zeta}]$$

obtemos,

$$E[\ln L(\boldsymbol{\zeta})] = \bar{C} + \sum_{i=1}^I \sum_{l=1}^q \{ \bar{r}_{il} \ln P_{il} + (\bar{f}_{il} - \bar{r}_{il}) \ln Q_{il} \}. \quad (2.29)$$

Dessa foram, os passos E e M são :

### Passo E

Usar os pontos de quadratura  $\bar{\theta}_l$ , os pesos  $A_l^{(t)}$ ,  $l = 1, \dots, q$  e as estimativas no instante  $t$  dos parâmetros dos itens,  $\hat{\boldsymbol{\zeta}}_i^{(t)}$ ,  $i = 1, \dots, I$ , para gerar  $g_j^*(\bar{\theta}_l)^{(t)}$  e, posteriormente,  $\bar{r}_{il}^{(t)}$  e  $\bar{f}_{il}^{(t)}$ ,  $i = 1, \dots, I$  e  $l = 1, \dots, q$ .

### Passo M

Com  $\mathbf{r}^{(t)}$  e  $\mathbf{f}^{(t)}$  obtidos no Passo E, resolver as equações de estimação para  $\boldsymbol{\zeta}_i$ ,  $i = 1, \dots, I$ , usando o algoritmo de Newton-Raphson ou Escore de Fisher.

Podemos notar que, a menos dos termos constantes, a expressão (2.29) é equivalente a (2.10), substituindo-se  $f_{il}$  e  $r_{il}$  por  $\bar{f}_{il}$  e  $\bar{r}_{il}$ , respectivamente. Dessa forma os algoritmos de maximização (Newton-Raphson/Escore de Fisher) podem ser implementados usando-se as expressões de (2.11) a (2.13), com as devidas substituições.

Contudo, Bock & Aitkin (1981) alertam que existe uma diferença conceitual entre esta extensão (adaptação) do princípio do algoritmo EM e o algoritmo EM geral de Dempster, Laird & Rubin (1977). Enquanto estes (para modelos fora da família exponencial) tomam o valor esperado da log-verossimilhança dos dados faltantes condicionados aos dados observados e as estimativas provisórias dos parâmetros, aqueles substituem os valores esperados dos dados faltantes (condicionados aos dados observados e as estimativas provisórias dos parâmetros) dentro da log-verossimilhança. Como visto, este procedimento satisfaz as equações de verossimilhança marginal e, além disso, reduz-se ao caso especial do algoritmo EM quando o modelo pertence à família exponencial e a log-verossimilhança é uma função linear de estatísticas suficientes [Bock & Aitkin(1981)].

Além disso, o uso do **algoritmo EM** garante a atualização de  $f_{il}$  e  $r_{il}$ , além de melhorar a qualidade e a garantia da convergência dos processos iterativos, ao contrário do que se usássemos somente os algoritmos de Newton-Raphson/Escore de Fisher. Maiores detalhes podem ser encontrados em Hsu, Ackerman & Fan (1999) e Woodruff & Hanson (1996).

## 2.3 Estimação das habilidades

Nesta seção apresentaremos a estimação por máxima verossimilhança das habilidades (**traços latentes**) dos indivíduos. Como mencionado anteriormente, basicamente podemos encontrar duas situações na prática : quando desconhecemos ambos os tipos de parâmetros (itens e habilidades), ou quando conhecemos no máximo um tipo. Nas seções anteriores desenvolvemos processos de estimação para os parâmetros dos itens, tanto na situação em que conhecemos as habilidades como também quando as desconhecemos. No que diz respeito as habilidades, o processo é construído de forma a considerar os parâmetros dos itens como conhecidos, ou seja, introduzimos as estimativas dos parâmetros dos itens (obtidos numa primeira etapa), digamos  $\hat{\zeta}$  na verossimilhança original  $L(\zeta, \theta)$  obtendo assim uma verossimilhança perfilada digamos  $L(\theta, \hat{\zeta})$ , a qual possui propriedades semelhantes as de uma verossimilhança genuína [Cordeiro (1992)]. Dessa forma para o desenvolvimento nesta seção, consideraremos  $L(\theta, \hat{\zeta}) \equiv L(\theta)$ . Vale ressaltar que o processo aqui apresentado é válido também quando conhecemos os verdadeiros valores dos parâmetros dos itens.

Pela independência entre as respostas de diferentes indivíduos (S1) e a independência local (S2), podemos escrever a log-verossimilhança como :

$$l(\theta) = \ln L(\theta) = \sum_{j=1}^n \sum_{i=1}^I \{y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln Q_{ij}\}. \quad (2.30)$$

Seguindo o desenvolvimento de Andrade, Tavares & Valle (2000), obtemos as seguintes expressões relativas aos processos de estimação

#### Vetor Escore

$$S(\theta_j) = \sum_{i=1}^I a_i(1 - c_i)(y_{ij} - P_{ij})W_{ij}, \quad (2.31)$$

#### Matriz Hessiana

$$H(\theta_j) = \sum_{i=1}^I (y_{ij} - P_{ij})W_{ij} \{H_{ij} - (y_{ij} - P_{ij})W_{ij}h_{ij}^2\},$$

#### Informação de Fisher

$$I(\theta_j) = \sum_{i=1}^I P_{ij}^* Q_{ij}^* W_{ij} h_{ij}^2,$$

com

$$h_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left( \frac{\partial P_{ij}}{\partial \theta_j} \right) = a_i(1 - c_i)$$

e

$$H_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left( \frac{\partial^2 P_{ij}}{\partial \theta_j^2} \right) = a_i^2(1 - c_i)(1 - 2P_{ij}^*).$$

Desse modo, considerando  $\hat{\theta}_j^{(t)}$  uma estimativa de  $\theta_j$  na iteração  $t$ , temos que os métodos de **Newton-Raphson** e **Escore de Fisher** podem ser escritos como

#### Newton-Raphson

$$\hat{\theta}_j^{(t+1)} = \hat{\theta}_j^{(t)} - [H(\hat{\theta}_j^{(t)})]^{-1} S(\hat{\theta}_j^{(t)}) \quad (2.32)$$



### Escore de Fisher

$$\hat{\theta}_j^{(t+1)} = \hat{\theta}_j^{(t)} + \left[ I \left( \hat{\theta}_j^{(t)} \right) \right]^{-1} S \left( \hat{\theta}_j^{(t)} \right) \quad (2.33)$$

$t = 0, 1, 2, \dots$

Detalhes dos valores iniciais dos processos iterativos podem ser encontrados em Andrade, Tavares & Valle (2000). Temos que, os EMV das habilidades convergem em distribuição para a normal com média igual ao verdadeiro valor das habilidades e variância igual a inversa da matriz Hessiana (Informação de Fisher) [Sen & Singer (1992), Houtink & Boomsma (1996), Chang & Stout (1993) e Chang (1996)]. Vemos também que a estimação é feita para cada indivíduo em separado e que podemos usar como “valores verdadeiros” dos parâmetros dos itens, estimativas baseadas em qualquer método.

## 2.4 Comentários adicionais

Muito embora o Método de Máxima Verossimilhança Marginal tenha possibilitado um enorme avanço no uso (ajustes) dos **Modelos de Resposta ao Item - MRI**, alguns problemas permanecem. Por exemplo quando todos os indivíduos respondem correta ou incorretamente um determinado item, não é possível de se obter as estimativas dos parâmetros do mesmo. Em contrapartida, quando um indivíduo responde correta ou incorretamente todos os itens, não é possível de se obter estimativas de sua habilidade [Baker (1992)].

Contudo podemos enumerar algumas boas propriedades do método de MVM [Baker (1992)]:

1. O processo de marginalização da verossimilhança com relação a  $\theta$  produz estimativas consistentes dos parâmetros dos itens, desde que o MRI e o modelo populacional (a priori) das habilidades sejam corretos, ou pelo menos se adequem de modo razoável.
2. A conceitualização dos indivíduos como uma amostra aleatória de uma população cuja as habilidades se distribuem segundo uma certa função densidade  $[g(\theta|\eta)]$ , permite assumir uma distribuição arbitrária para as mesmas na população amostrada. Conseqüentemente, a métrica das estimativas dos parâmetros dos itens é definida pelos parâmetros de localização e escala em  $\eta$ .
3. Permite a imposição de uma estrutura parecida com o procedimento bayesiano, *Bayesian-like*, de tal forma que a inferência a respeito as habilidades seja aperfeiçoada. Como visto anteriormente, o algoritmo EM apresentado não envolve a solução bayesiana clássica haja vista que os parâmetros dos itens são tratados como constantes.
4. A estrutura de programação requerida para a implementação computacional do método é relativamente simples apesar de todo o desenvolvimento matemático. Os passos do algoritmo podem ser bem organizados e seqüenciados.

Apesar de, por exemplo, para o MLUP1 existirem processos de estimação de Máxima Verossimilhança Condicional (MVCD), Thissen (1982), sugere que a acurácia e o tempo computacional são melhores via MVM do que MVCD. Enquanto que naquele o incremento no tempo de execução é da ordem de  $I$ , para esse o incremento é da ordem de  $I^2$ , com relação ao número de itens.

---

## Capítulo 3

# Modelos Logísticos Unidimensionais para Respostas Dicotômicas: Estimação Bayesiana

---

### 3.1 Introdução

Neste Capítulo trataremos da estimação Bayesiana dos parâmetros do MLUP3.

Como vimos no Capítulo anterior, os métodos de Máxima Verossimilhança apresentam problemas com escores nulos e perfeitos (tanto para os itens quanto para os indivíduos), e também com padrões de resposta aberrantes (indivíduos com habilidades elevadas que respondem incorretamente itens fáceis, e vice-versa, indivíduos com habilidade baixa que respondem corretamente itens difíceis).

Basicamente, os pesquisadores na TRI usam métodos Bayesianos para tentar aprimorar a acurácia das estimativas dos parâmetros dos modelos propostos e contornar os problemas existentes nos EMV, sem se preocupar com as nuances filosóficas da Inferência Bayesiana [Baker(1992)]. Notadamente, as estimativas de máxima verossimilhança do parâmetro de acerto casual podem se apresentar por demais discrepantes. Isto ocorre com certa freqüência na prática, principalmente quando dispomos de poucos indivíduos com níveis de habilidade baixos. Os métodos bayesianos conseguiram melhorar em muito o processo de estimação desse parâmetro [Baker(1992)]. Com esse espírito, algumas idéias foram propostas nos últimos tempos.

Ante a situação em que desconhecemos apenas um tipo de parâmetro, e desejamos estimar o outro (em situações como as descritas no Capítulo 2), os procedimentos Bayesianos convencionais podem ser usados, como será visto na próxima Seção. Entretanto, em situações de desconhecimento de ambos os tipos, abordagens específicas tornam-se necessárias.

Nesse sentido Swaminathan & Gifford (1982, 1985, 1986) propuseram uma estimação Bayesiana conjunta dos parâmetros dos modelos MLUP1, MLUP2 e MLUP3, respectivamente, utilizando uma extensão do paradigma (abordagem) de Birnbaum. À semelhança desse procedimento clássico, a estimação das habilidades e dos parâmetros dos itens é executada simultaneamente. Em contrapartida, Mislavy (1986) e Tsutakawa & Lin (1986), propuseram métodos de estimação bayesiana que são extensões diretas do método de MVM de Bock & Aitkin (1981). Notadamente, a abordagem de Mislavy inclui as propriedades inerentes dos estimadores via MVM, porém tenta restringir as estimativas dos parâmetros (dos itens) para que não se obtenha valores discrepantes (como por exemplo, valores negativos para o parâmetro de discriminação). Além disso, existem

evidências empíricas que as soluções marginalizadas são superiores às ditas não-marginalizadas (conjuntas)[O'Hagan (1976)]. Com efeito, esta abordagem é a que será descrita para a estimação dos parâmetros dos itens ante ao desconhecimento das habilidades dos indivíduos.

Como a estimação das habilidades não envolve o problema da presença de parâmetros de perturbação, procedimentos bayesianos clássicos podem ser usados sem grandes adaptações [Baker (1992)].

Introduções gerais e discussões conceituais a respeito da metodologia da Inferência Bayesiana podem ser encontradas em Cornfield (1969), de Finnetti (1974), Edwards, Lindman & Savage (1963), Lindley (1970a, 1970b), Novick & Jackson (1974) e Bernardo & Smith (1994), por exemplo.

Na próxima Seção abordaremos a estimação dos parâmetros dos itens.

## 3.2 Estimação dos parâmetros dos itens

Esta seção trata da estimação Bayesiana dos parâmetros dos itens no MLUP3. Novamente, para os outros dois modelos (de 1 e 2 parâmetros) as expressões podem ser obtidas como casos particulares do desenvolvimento aqui descrito. As notações são basicamente as mesmas do Capítulo 2, introduzindo-se algumas outras.

### 3.2.1 Habilidades conhecidas

Nessa parte, desenvolveremos o processo de estimação bayesiana para os parâmetros dos itens considerando as habilidades conhecidas. Basicamente as estimativas bayesianas são construídas a partir da distribuição a posteriori, que une a informação contida na verossimilhança com a distribuição a priori. Pelo **Teorema de Bayes** podemos construir a distribuição a posteriori da seguinte forma :

$$\begin{aligned}
 f(\zeta | \mathbf{y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau}) &= \frac{f(\zeta, \mathbf{Y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau})}{\int_{\mathbb{R}^q} f(\zeta, \mathbf{Y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau}) d\zeta} \propto f(\mathbf{Y}_{..} | \boldsymbol{\theta}, \zeta, \boldsymbol{\tau}) f(\zeta, \boldsymbol{\tau}) \\
 &= P(\mathbf{Y}_{..} = \mathbf{y}_{..} | \boldsymbol{\theta}, \zeta) f(\zeta | \boldsymbol{\tau}) f(\boldsymbol{\tau}) \\
 &= \left\{ \prod_{i=1}^I \prod_{j=1}^n P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}} \right\} f(\zeta | \boldsymbol{\tau}) f(\boldsymbol{\tau}). \tag{3.1}
 \end{aligned}$$

com  $\boldsymbol{\tau} = (\boldsymbol{\tau}'_1, \dots, \boldsymbol{\tau}'_I)'$ , eventualmente  $\boldsymbol{\tau}_i = (\boldsymbol{\tau}'_{a_i}, \boldsymbol{\tau}'_{b_i}, \boldsymbol{\tau}'_{c_i})' \equiv \boldsymbol{\tau}_i^* = (\boldsymbol{\tau}_a^*, \boldsymbol{\tau}_b^*, \boldsymbol{\tau}_c^*)'$ , representando os **hiperparâmetros** relacionados a  $\zeta_i$  e  $q$  o número de parâmetros dos itens.

Notemos que o primeiro termo de (3.1) nada mais é que a verossimilhança descrita em (2.3), a segunda é uma distribuição a priori para os parâmetros do itens e o terceiro termo é uma distribuição associada aos **hiperparâmetros** que caracterizam a distribuição a priori do vetor

$\zeta$ . A especificação da distribuição a priori para os parâmetros de interesse e da distribuição dos hiperparâmetros constituem, respectivamente, o primeiro e o segundo estágio do **modelo bayesiano (hierárquico)** proposto por Lindley & Smith (1972). Muito embora esta última etapa possa ser ignorada, haja vista que, na prática, em problemas de estimação na TRI, são atribuídos valores para o vetor  $\tau$  na distribuição a priori de  $\zeta$ , iremos mantê-la para fins de se obter uma abordagem mais geral. Apesar de muitas quantidades da distribuição a priori (média, moda, mediana etc) possam ser escolhidas como estimadores dos parâmetros, desenvolveremos o processo baseado na moda a posteriori (MAP), haja vista ser esse procedimento mais difundido e adotado na estimação bayesiana dos parâmetros dos itens [Baker (1992)].

Como não há razão para acreditar que os parâmetros de um determinado item tenham informações a respeito dos parâmetros de outros, é bastante razoável atribuir prioris independentes para  $\zeta$ , ou seja, considerar os parâmetros de diferentes itens como **estocasticamente independentes**. Dessa forma podemos reescrever (3.1) como o seguinte:

$$f(\zeta|\mathbf{y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau}) \propto \left\{ \prod_{i=1}^I \prod_{j=1}^n P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}} \right\} \left\{ \prod_{i=1}^I f(\zeta_i|\tau_i) \right\} \left\{ \prod_{i=1}^I f(\tau_i) \right\}. \quad (3.2)$$

Tomando o logaritmo natural de (3.2), temos que

$$\ln f(\zeta|\mathbf{y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau}) \propto l(\zeta) + \sum_{i=1}^I \ln f(\zeta_i|\tau_i) + \sum_{i=1}^I \ln f(\tau_i), \quad (3.3)$$

com  $l(\zeta) = \sum_{i=1}^I \sum_{j=1}^n \{y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln Q_{ij}\}$ . Como vemos, a distribuição a posteriori é contínua (devido ao espaço paramétrico associado ao modelo) e dessa forma a moda é o conjunto de valores que maximizam (3.1). Portanto, derivando equivalentemente (3.3) com respeito a  $\zeta_i$  obteremos o vetor de **funções de estimação bayesianas**

$$\mathcal{S}(\zeta_i)_B = \frac{\partial \ln f(\zeta|\mathbf{y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau})}{\partial \zeta_i} = \frac{\partial l(\zeta)}{\partial \zeta_i} + \frac{\partial \ln f(\zeta_i|\tau_i)}{\partial \zeta_i}. \quad (3.4)$$

Notamos que a primeira parcela do lado direito de (3.4) gera as equações de estimação de máxima verossimilhança (2.4). Sendo assim desenvolveremos apenas o segundo termo, que é devido as prioris. Como cada parâmetro tem características próprias abordaremos separadamente suas respectivas distribuições.

#### Distribuição a priori para $a_i$

Geralmente, adota-se as distribuições *Log-normal* ou *Qui-Quadrado* para  $a_i$ . Neste trabalho, vamos considerar que cada parâmetro  $a_i$  tem distribuição Log-normal com parâmetro  $\tau_{a_i} = (\mu_{a_i}, \sigma_{a_i}^2)'$ . Uma justificativa teórica para a adoção desta distribuição é que na prática os

$a_i$  são, em geral, positivos (o modelo pressupõe esse fato), sugerindo que a distribuição de  $a_i$  pode ser modelada por uma distribuição unimodal e com assimetria positiva [veja Mislevy (1986)], tal como a log-normal. A transformação  $\alpha_i = \log a_i$  resulta em cada  $\alpha_i$  tendo uma distribuição normal  $(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)$ , onde  $\mu_{\alpha_i} = \exp[\mu_{a_i} + \sigma_{a_i}^2/2]$  e  $\sigma_{\alpha_i}^2 = [\exp(\sigma_{a_i}^2) - 1] \exp[2\mu_{\alpha_i} + \sigma_{\alpha_i}^2]$ . Alguns autores (Baker 1992, por exemplo) preferem desenvolver expressões para a estimação de  $\alpha_i$  ao invés de  $a_i$  e sugerem a utilização da propriedade de invariância do estimador de máxima verossimilhança para a obtenção de  $\hat{a}_i$  pela transformação  $\hat{a}_i = \exp(\hat{\alpha}_i)$ . Entretanto, para uniformidade desse trabalho, vamos continuar apresentando as equações para o parâmetro  $a_i$ .

Como a distribuição de  $a_i$  é log-normal, sua densidade é

$$f(a_i|\mu_{a_i}, \sigma_{a_i}^2) = \frac{1}{\sqrt{2\pi}a_i\sigma_{a_i}} \exp\left[-\frac{1}{2\sigma_{a_i}^2}(\ln a_i - \mu_{a_i})^2\right].$$

Segue que a segunda parcela de (3.4) pode ser escrita como

$$\frac{\partial \ln f(a_i|\mu_{a_i}, \sigma_{a_i}^2)}{\partial a_i} = -\frac{1}{a_i} \left[1 + \frac{\ln a_i - \mu_{a_i}}{\sigma_{a_i}^2}\right]. \quad (3.5)$$

### Distribuição a priori para $b_i$

Vamos considerar que cada  $b_i$  tem distribuição *Normal* com vetor de parâmetros  $\boldsymbol{\tau}_{b_i} = (\mu_{b_i}, \sigma_{b_i}^2)'$ . Isso é bastante razoável, uma vez que este parâmetro pode ser visto como uma “habilidade” do item, e é medido na mesma escala que as habilidades, e além disso, pode assumir qualquer valor na reta. Desta forma, a segunda parcela de (3.4) pode ser escrita como

$$\frac{\partial \ln f(b_i|\mu_b, \sigma_b^2)}{\partial b_i} = -\frac{(b_i - \mu_b)}{\sigma_b^2}. \quad (3.6)$$

### Distribuição a priori para $c_i$

Como  $c_i$  é uma probabilidade, seu valor deve pertencer ao intervalo  $[0; 1]$  e, em geral, utiliza-se uma priori *Beta*, conforme proposto por Swaminathan & Gifford (1986). A densidade da distribuição Beta  $(\alpha_i - 1, \beta_i - 1)$  é dada por

$$f(c_i|\alpha_i, \beta_i) = \frac{\Gamma(\alpha_i + \beta_i - 2)}{\Gamma(\alpha_i - 1)\Gamma(\beta_i - 1)} c_i^{\alpha_i-2} (1 - c_i)^{\beta_i-2}.$$

Para chegarmos à expressão para a segunda parcela de (3.4), notemos que

$$\ln f(c_i|\alpha_i, \beta_i) = \text{Const} + (\alpha_i - 2) \ln c_i + (\beta_i - 2) \ln(1 - c_i).$$

Consequentemente,

$$\frac{\partial \ln f(c_i | \alpha_i, \beta_i)}{\partial c_i} = \frac{\alpha_i - 2}{c_i} - \frac{\beta_i - 2}{1 - c_i}. \quad (3.7)$$

O uso da distribuição Beta como priori, nos leva a interpretar  $p_i$  (valor esperado da distribuição) como a probabilidade de acerto de indivíduos com baixa habilidade, i. e., um valor *a priori* para  $c_i$ . A idéia é especificar os valores de  $\alpha_i$  e  $\beta_i$  de modo que se tenha o valor de  $p_i$  desejado. Em geral, atribui-se,  $p_i = 1/nalt_i$ , com  $nalt_i$  sendo o número de alternativas do item  $i$ . Além disso, os valores propostos para  $m_i$  também influenciam os intervalos de credibilidade (semelhantes aos intervalos de confiança clássicos) construídos para  $c_i$  [Swaminathan & Gifford (1986) e Novick & Jackson (1974), pg. 119]. Estes intervalos levam em consideração a priori e fornecem uma medida de credibilidade de que o parâmetro pertença a um determinado intervalo. Além disso, pesos maiores das prioris (variâncias menores) levam a intervalos mais estreitos. Isto sugere que um intervalo de credibilidade possa ser usado em vez de se especificar a variância da distribuição a priori [Novick & Jackson (1974)]. Entretanto, Swaminathan & Gifford sugerem que a escolha de  $m_i$  deva se situar no intervalo  $\{15, \dots, 20\}$ , o que leva a uma certa restrição na escolha de  $\alpha_i$  e  $\beta_i$ .

Dessa forma, usando (2.4) e (3.5) a (3.7), temos que o vetor escore das equações de estimação bayesianas (EEB), é dado por

$$\mathbf{S}(\boldsymbol{\zeta}_i)_B = \sum_{j=1}^n (y_{ij} - P_{ij}) W_{ij} \mathbf{h}_{ij} + \boldsymbol{\lambda}_i, \quad (3.8)$$

com

$$\boldsymbol{\lambda}_i = \left[ \frac{1}{a_i} \left[ 1 + \frac{\ln a_i - \mu_{a_i}}{\sigma_{a_i}^2} \right]; -\frac{(b_i - \mu_{b_i})}{\sigma_{b_i}^2}; \frac{\alpha_i - 2}{c_i} - \frac{\beta_i - 2}{1 - c_i} \right]'. \quad (3.9)$$

Observemos que o sistema de equações descrito em (3.8) é não-linear o que nos leva a ter que usar algum método iterativo (Newton-Raphson/Escore de Fisher). Para tanto precisaremos da Matriz Hessiana e da Informação de Fisher, respectivamente.

Notemos que ao se derivar (3.8) com respeito a  $\boldsymbol{\zeta}_i$  teremos duas parcelas. A primeira é a Matriz Hessiana referente ao processo de máxima verossimilhança e a segunda são as derivadas cruzadas com relação as prioris. Notando que as distribuições adotadas são independentes, temos que estas derivadas são dadas por

$$\frac{\partial^2 f(a_i | \mu_{a_i}, \sigma_{a_i}^2)}{\partial a_i^2} = \frac{[\sigma_{a_i}^2 + \ln a_i - \mu_{a_i} - 1]}{\mu_{a_i}^2 \sigma_{a_i}^2} \quad (3.10)$$

$$\frac{\partial^2 f(b_i | \mu_{b_i}, \sigma_{b_i}^2)}{\partial b_i^2} = -\frac{1}{\sigma_{b_i}^2} \quad (3.11)$$

$$\frac{\partial^2 f(c_i | \alpha_i, \beta_i)}{\partial c_i^2} = -\left[ \frac{\alpha_i - 2}{c_i^2} + \frac{\beta_i - 2}{(1 - c_i)^2} \right], \quad (3.12)$$

e as derivadas cruzadas são todas nulas. Portanto de (3.8), (3.10) a (3.12) e usando (2.5), temos que a Matriz Hessiana é dada por

$$\mathbf{H}(\zeta_i)_B = \sum_{j=1}^n (y_{ij} - P_{ij}) W_{ij} \left\{ \mathbf{H}_{ij} - W_{ij} (y_{ij} - P_{ij}) \mathbf{h}_{ij} \mathbf{h}'_{ij} \right\} + \mathbf{\Lambda}_i, \quad (3.13)$$

onde

$$\mathbf{\Lambda}_i = \begin{bmatrix} \frac{[\sigma_{a_i}^2 + \ln a_i - \mu_{a_i} - 1]}{\mu_{a_i}^2 \sigma_{a_i}^2} & \cdot & \cdot \\ 0 & -\frac{1}{\sigma_{b_i}^2} & \cdot \\ 0 & 0 & -\left[ \frac{\alpha_i - 2}{c_i^2} \right] + \frac{\beta_i - 2}{(1 - c_i)^2} \end{bmatrix}. \quad (3.14)$$

Para o cálculo da Informação de Fisher, lembremos que  $\mathbf{\Lambda}_i$  é uma matriz não-estocástica (com relação as respostas), o que unido a (2.6) resulta em

$$\mathbf{I}(\zeta_i)_B = \mathbb{E} \{ \mathbf{H}(\zeta_i)_B | \zeta_i \} = \sum_{j=1}^n P_{ij}^* Q_{ij}^* W_{ij} \mathbf{h}_{ij} \mathbf{h}'_{ij} - \mathbf{\Lambda}_i. \quad (3.15)$$

### Habilidades conhecidas e agrupadas

Considerando o agrupamento das habilidades descrito no Capítulo 2, não é difícil verificar que as expressões calculadas anteriormente se transformam em

$$\mathbf{S}(\zeta_i)_B = \sum_{l=1}^q \{r_{il} - f_{il} P_{il}\} W_{il} \mathbf{h}_{il} + \boldsymbol{\lambda}_i, \quad (3.16)$$

$$\mathbf{H}(\zeta_i)_B = \sum_{l=1}^q \left[ W_{il} \left\{ (r_{il} - f_{il} P_{il}) \mathbf{H}_{il} - W_{il} (r_{il} - 2r_{il} P_{il} + f_{il} P_{il}^2) \mathbf{h}_{il} \mathbf{h}'_{il} \right\} \right] + \mathbf{\Lambda}_i \quad (3.17)$$



e

$$\mathbf{I}(\zeta_i)_B = \sum_{l=1}^q f_{il} P_{il}^* Q_{il}^* \mathbf{h}_{il} \mathbf{h}'_{il} - \Lambda_i. \quad (3.18)$$

Na próxima Seção apresentaremos o processo de estimação ante o desconhecimento das habilidades.

### 3.2.2 Habilidades desconhecidas

Como já foi dito, no que se refere a estimação dos parâmetros dos itens ante o desconhecimento das habilidades, apresentaremos o método de **Estimação Bayesiana Marginal** (EBM) proposto por Mislevy (1986). Basicamente, a EBM é uma extensão direta do método de MVM, na medida em que se baseia no mesmo processo de construção da **verossimilhança marginal**, acoplado à mesma, prioris convenientes (relativas aos parâmetros dos itens). Subseqüentemente, já então de posse de uma distribuição a posteriori (marginal) utiliza-se a mesma idéia do algoritmo EM, descrita na Seção 2.2.2, de modo a maximizar a referida posteriori.

Devido a essa idéia, o presente método também é conhecido por MMAP (*marginal maximum a posteriori*). Por se tratar de uma extensão de método de MVM, muitas das notações e/ou expressões serão semelhantes, senão iguais àquelas desenvolvidas no Capítulo 2. Portanto, o desenvolvimento será focado nos termos novos.

Sendo assim, consideremos primeiramente o vetor de parâmetros  $\psi = (\theta', \zeta', \eta', \tau')'$ . Sendo que  $\eta$  e  $\tau$  representam, respectivamente, os **hiperparâmetros** relativos a  $\theta$  e  $\zeta$ . Do mesmo modo que na Seção 3.2.1, sob a hipótese de independência estocástica entre os parâmetros, temos que a distribuição a posteriori é dada por

$$\begin{aligned} f(\psi|\mathbf{y}_{..}) &= CL(\mathbf{y}_{..}; \theta, \zeta) f(\zeta|\tau) g(\theta|\eta) f(\tau) g(\eta) \\ &\propto L(\mathbf{y}_{..}; \theta, \zeta) f(\zeta|\tau) g(\theta|\eta) f(\tau) g(\eta), \end{aligned} \quad (3.19)$$

onde  $C$  representa uma constante de normalização,  $L(\mathbf{y}_{..}; \theta, \zeta)$  a verossimilhança genuína associada ao nosso problema,  $f(\zeta|\tau)$  e  $g(\theta|\eta)$  as prioris associadas a  $\zeta$  e  $\theta$ , respectivamente,  $f(\tau)$  e  $g(\eta)$  as **hiperprioris** associadas a  $\tau$  e  $\eta$ , respectivamente. Seguindo Bock & Lieberman (1970), devemos integrar (3.19) com relação as habilidades, eliminado-se assim, o problema de estimá-las junto com os parâmetros dos itens. Como vimos na Seção 2.2.2, se o **MRI** e a **distribuição das habilidades** forem corretamente especificadas, este processo leva a estimativas consistentes dos parâmetros dos itens (para testes de comprimento finito) com o aumento do número de indivíduos.

No entanto, Mislevy (1986) sugere que em muitas situações práticas o vetor  $\tau$  não é de

interesse, podendo ser tratado, também, como parâmetro de perturbação. Em contrapartida, o parâmetro  $\eta$ , pode ser de interesse. Dessa forma, o manteremos na distribuição a posteriori.

Portanto, o vetor de parâmetros de perturbação passa a ser  $\phi = (\theta', \tau)'$  e, assim integrando (3.19) com relação a  $\phi$  vem que

$$\begin{aligned} f(\psi|\mathbf{y}_{..}) &= C \int \int L(\mathbf{y}_{..}; \theta, \zeta) f(\zeta|\tau) g(\theta|\eta) f(\tau) g(\eta) d\theta d\tau \\ &= C g(\eta) \left\{ \int L(\mathbf{y}_{..}; \theta, \zeta) g(\theta|\eta) d\theta \right\} \left\{ \int f(\zeta|\tau) f(\tau) d\tau \right\} \\ &\propto L(\zeta, \eta) f(\zeta) g(\eta), \end{aligned} \quad (3.20)$$

com  $L(\zeta, \eta) \equiv P(\mathbf{Y}_{..} = \mathbf{y}_{..} | \zeta, \eta)$ ,  $f(\zeta) \equiv \int f(\zeta|\tau) f(\tau) d\tau$  e  $\psi = (\zeta', \eta)'$ .

Notemos que, muito embora tenhamos eliminado  $\theta$  de (3.20), isso não retira a influência de  $\eta$ , ou seja, temos que especificar essa componente (o que garante a determinação da métrica das estimativas). O mesmo pode ser dito a respeito do vetor  $\tau$ , o qual mesmo tendo sido integrado de (3.20), ainda assim, faz-se necessário sua especificação na distribuição  $f(\zeta)$  [Baker(1992)].

Dessa forma, o logaritmo natural de (3.20) é dado por

$$\begin{aligned} \ln f(\psi|\mathbf{y}_{..}) \equiv l(\psi|\mathbf{y}_{..}) &\propto \ln L(\zeta, \eta) + \ln f(\zeta) + \ln g(\eta) \\ &\equiv l(\zeta, \eta) + \ln f(\zeta) + \ln g(\eta). \end{aligned} \quad (3.21)$$

Como o interesse reside no máximo da posteriori (moda), devemos derivar (3.21) com relação a  $\zeta_i$ , o que resulta em

$$\mathbf{S}(\zeta_i)_{BM} = \frac{\partial l(\zeta, \eta)}{\partial \zeta_i} + \frac{\partial \ln f(\zeta)}{\partial \zeta_i}. \quad (3.22)$$

Notemos que a primeira parcela é igual àquela desenvolvida na Seção 2.2.2 e a segunda é igual àquela apresentada em (3.9). Dessa forma de (2.11) e (3.9) temos que

$$\mathbf{S}(\zeta_i)_{BM} = \sum_{l=1}^q (\bar{r}_{il} - \bar{f}_{il} P_{il}) W_{il} \mathbf{h}_{il} + \lambda_i. \quad (3.23)$$

Novamente para aplicação dos métodos iterativos precisaremos da matriz Hessiana e da informação de Fisher. De (2.12), (2.13) e (3.14) em (3.22), temos que essas componentes são dadas respectivamente por

$$\mathbf{H}(\zeta_i)_{BM} = \sum_{l=1}^q \left[ W_{il} \left\{ (\bar{r}_{il} - \bar{f}_{il} P_{il}) \mathbf{H}_{il} - W_{il} (\bar{r}_{il} - 2\bar{r}_{il} P_{il} + \bar{f}_{il} P_{il}^2) \mathbf{h}_{il} \mathbf{h}'_{il} \right\} \right] + \Lambda_i \quad (3.24)$$

e

$$I(\zeta_i)_{BM} = \sum_{l=1}^q \bar{f}_{il} P_{il}^* Q_{il}^* \mathbf{h}_{il} \mathbf{h}'_{il} - \Lambda_i. \quad (3.25)$$

Dessa forma, o algoritmo EM pode ser implementado como descrito na Seção 2.2.2, usando-se a esperança da log-posteriori (3.21) ao invés da esperança da verossimilhança.

Vemos que, basicamente, o método de EBM adiciona termos nas expressões obtidas via MVM, relativas as prioris. Na próxima Seção apresentaremos a estimação das habilidades.

### 3.3 Estimação das habilidades

Da mesma forma que na estimação por MV, a estimação bayesiana das habilidades é realizada numa segunda etapa, considerando conhecidos os parâmetros dos itens.

Considerando que indivíduos diferentes não possuem informação de outros indivíduos, podemos considerar que seus traços latentes são **estocasticamente independentes**, e, dessa forma, tratar os respectivos processos de estimação em separado. Portanto, a distribuição a posteriori do traço latente de um determinado indivíduo é dada por

$$g_j^*(\theta_j) \equiv Cg(\theta_j | \mathbf{y}_{.j}, \zeta, \boldsymbol{\eta}) = C P(\mathbf{Y}_{.j} | \theta_j, \zeta) g(\theta_j | \boldsymbol{\eta}) \quad (3.26)$$

$$\propto P(\mathbf{Y}_{.j} | \theta_j, \zeta) g(\theta_j | \boldsymbol{\eta}) \quad (3.27)$$

Freqüentemente, na literatura adota-se a média [conhecido como EAP (*expectation a posteriori*)], ou a moda [conhecido com MAP (*maximum a posteriori*)] como estimadores das habilidades. Consideremos primeiramente a moda.

#### Estimação pela moda a posteriori - MAP

Para calcular a moda, devemos encontrar o máximo da distribuição a posteriori (3.27). Convenientemente, tomando o logaritmo natural de (3.27), vem que

$$\ln g_j^*(\theta_j) \equiv l_j^*(\theta_j) \propto l(\theta_j) + \ln g(\theta_j | \boldsymbol{\eta}), \quad (3.28)$$

com  $l(\theta_j) = \prod_{i=1}^I P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}}$ . Segue de (3.28), que a função de estimação bayesiana, para uma determinada habilidade, é dada por

$$S(\theta_j)_B = \frac{\partial l_j^*(\theta_j)}{\partial \theta_j} = \frac{\partial l(\theta_j)}{\partial \theta_j} + \frac{\partial \ln g(\theta_j | \boldsymbol{\eta})}{\partial \theta_j}. \quad (3.29)$$

Notemos que, o primeiro termo de (3.29) é igual ao obtido em (2.31), e assumindo distribuição Normal com  $\eta = (\mu_\theta, \sigma_\theta^2)'$ , o segundo termo é dado por

$$\frac{\partial \ln g(\theta_j, \boldsymbol{\eta})}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \left\{ \text{const} - \frac{(\theta_j - \mu_\theta)^2}{2\sigma_\theta^2} \right\} = -\frac{\theta_j - \mu_\theta}{\sigma_\theta^2}, \quad (3.30)$$

portanto de (2.31) e (3.30), temos que

$$S(\theta_j)_B = \sum_{i=1}^I a_i (1 - c_i) (y_{ij} - P_{ij}) - \frac{\theta_j - \mu_\theta}{\sigma_\theta^2}. \quad (3.31)$$

Novamente, pela não linearidade da equação de estimação, devemos usar algum método iterativo para resolvê-la. Usando o fato de que

$$\frac{\partial^2 \ln g(\theta_j | \boldsymbol{\eta})}{\partial \theta_j^2} = -\frac{1}{\sigma_\theta^2}, \quad (3.32)$$

temos que, a matriz Hessiana e a informação de Fisher são dadas respectivamente por

$$H(\theta_j)_B = \sum_{i=1}^I (y_{ij} - P_{ij}) W_{ij} \{ H_{ij} - (y_{ij} - P_{ij}) W_{ij} h_{ij}^2 \} - \frac{1}{\sigma_\theta^2} \quad (3.33)$$

e

$$I(\theta_j)_B = \sum_{i=1}^I P_{ij}^* Q_{ij}^* W_{ij} h_{ij}^2 + \frac{1}{\sigma_\theta^2}. \quad (3.34)$$

Dessa forma os métodos iterativos podem ser implementados tais como descritos em (2.32) e (2.33). O procedimento de estimação MAP sempre converge, independentemente do padrão de respostas do indivíduo [Mislevy & Bock (1989)]. Consideraremos agora a média da distribuição posteriori.

### Estimação pela média a posteriori - EAP

De (3.26), temos que a esperança a posteriori é dada por

$$\mathbb{E}[\theta_j | \mathbf{y}_j, \boldsymbol{\zeta}, \boldsymbol{\eta}] = \frac{\int_{\mathbb{R}} \theta P(\mathbf{Y}_j | \theta, \boldsymbol{\zeta}) g(\theta | \boldsymbol{\eta}) d\theta}{\int_{\mathbb{R}} P(\mathbf{Y}_j | \theta, \boldsymbol{\zeta}) g(\theta | \boldsymbol{\eta}) d\theta}. \quad (3.35)$$

Podemos notar que as integrais em (3.35) não possuem solução explícita, portanto devemos usar algum procedimento numérico. Utilizando o método da *Quadratura Gaussiana* descrito na Seção 2.2.2, temos que (3.35) se transforma em

$$\begin{aligned} \mathbb{E} [\theta_j | \mathbf{y}_{.j}, \boldsymbol{\zeta}, \boldsymbol{\eta}] &\approx \mathbb{E} [\bar{\theta}_j | \mathbf{y}_{.j}, \boldsymbol{\zeta}, \boldsymbol{\eta}] = \frac{\sum_{l=1}^q \bar{\theta}_l P(\mathbf{Y}_{.j} | \bar{\theta}_l, \boldsymbol{\zeta}) g(\bar{\theta}_l | \boldsymbol{\eta})}{\sum_{l=1}^q P(\mathbf{Y}_{.j} | \bar{\theta}_l, \boldsymbol{\zeta}) g(\bar{\theta}_l | \boldsymbol{\eta})} \\ &= \frac{\sum_{l=1}^q \bar{\theta}_l P(\mathbf{Y}_{.j} | \bar{\theta}_l, \boldsymbol{\zeta}) A_l}{\sum_{l=1}^q P(\mathbf{Y}_{.j} | \bar{\theta}_l, \boldsymbol{\zeta}) A_l}. \end{aligned} \quad (3.36)$$

A expressão (3.36) tem algumas características interessantes. Além de ter a vantagem de ser calculada diretamente, ou seja, sem a necessidade de métodos iterativos, as quantidades necessárias para seu cálculo são um produto final da etapa de estimação dos parâmetros dos itens. Por isso, Mislevy & Stocking (1989), recomendam seu uso para a estimação das habilidades.

Uma medida de precisão do EAP pode ser obtida via variância a posteriori

$$\text{Var} [\theta_j | \mathbf{y}_{.j}, \boldsymbol{\zeta}, \boldsymbol{\eta}] = \frac{\int_{\mathbb{R}} [\theta - \mathbb{E}(\theta)]^2 P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\zeta}) g(\theta | \boldsymbol{\eta}) d\theta}{\int_{\mathbb{R}} P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\zeta}) g(\theta | \boldsymbol{\eta}) d\theta}$$

Em termos dos pontos de quadratura, temos que

$$\text{Var} [\bar{\theta}_j | \mathbf{y}_{.j}, \boldsymbol{\zeta}, \boldsymbol{\eta}] = \frac{\sum_{l=1}^q \{\bar{\theta}_l - \mathbb{E}[\bar{\theta}_j | \mathbf{y}_{.j}, \boldsymbol{\zeta}, \boldsymbol{\eta}]\}^2 P(\mathbf{Y}_{.j} | \bar{\theta}_l, \boldsymbol{\zeta}) g(\bar{\theta}_l | \boldsymbol{\eta})}{\sum_{l=1}^q P(\mathbf{Y}_{.j} | \bar{\theta}_l, \boldsymbol{\zeta}) g(\bar{\theta}_l | \boldsymbol{\eta})} \quad (3.37)$$

Para a escolha dos pesos  $A_l$ , podemos usar **pesos teóricos** (assumindo uma distribuição teórica para as habilidades), **pesos empíricos** (estimados durante a calibração dos itens) [Mislevy (1984)] ou **pesos subjetivos** (fornecidos pelo pesquisador) [Mislevy & Bock (1990)].

### 3.4 Comentários adicionais

Os métodos de estimação bayesiana na TRI representam um grande avanço no processo de implementação dos MRI, uma vez que permitem a obtenção de estimativas mesmo na presença de padrões de resposta aberrantes, e ainda, conseguiram diminuir o vício das estimativas.

No entanto surgem outros problemas. Por exemplo, Baker (1992) adverte que a má escolha das priors pode levar a resultados viesados. Mais especificamente, *prioris* com variâncias pequenas (informativas) e médias distantes dos verdadeiros valores dos parâmetros dos itens podem levar a resultados totalmente distorcidos.

---

Harwell & Janosky (1991), baseados em resultados empíricos, sugerem que as variâncias das *prioris* para os parâmetros dos itens têm pouco efeito nas estimativas dos mesmos, quando o número de indivíduos é maior que 250. No entanto, os resultados de Baker (1990), indicam que, mesmo para um grande número de indivíduos, as *prioris* especificadas erroneamente, podem introduzir o que ele chamou de **vícios conjuntos** nos resultados [Baker (1992)]. Em alguns casos, estimar os **hiperparâmetros** com base nos dados, e atualizá-los em cada passo dos processos iterativos, pode ser efetivo na prevenção destes vícios [Baker (1992) e Mislevy & Bock (1990)]. Uma outra alternativa seria usar as estimativas de MVM para se ter uma idéia dos valores dos parâmetros.

---

## Capítulo 4

# Modelo de Resposta Nominal: Apresentação e Estimação por Máxima Verossimilhança

---

### 4.1 Introdução

Os itens de respostas policotômicas são basicamente aqueles que possuem mais de duas categorias de resposta (**não ordenadas** ou **nominais**), ou itens dissertativos (abertos), para o qual são atribuídos escores numa determinada escala (**categorias ordenadas** ou **ordinais**), por exemplo uma nota atribuída no conjunto dos números inteiros positivos.

O início do desenvolvimento dos **Modelos para itens Policotômicos** se deu no final dos anos 60, mas só começou a se tornar realmente assunto de pesquisas no início dos anos 80 [Van der Linden & Hambleton (1997)].

Bock (1972), propôs o chamado **Modelo de Resposta Nominal**, o qual objetivava analisar itens com categorias não ordenadas (tais como em testes de múltipla escolha) e deles poder extrair mais informação a respeito das habilidades (proficiência) dos indivíduos do que simplesmente considerar que os mesmos tenham respondido ou não corretamente a estes itens.

Por outro lado, Samejima (1969) introduziu um modelo muito importante conhecido por **Modelo de Resposta Gradual**, para analisar dados oriundos de escalas de atitudes de Likert e de performances em questões corrigidas de maneira gradual [van der Linden & Hambleton(1997)]. Este modelo e suas variantes foram desenvolvidas para analisar itens de categorias ordenadas. Estes trabalhos foram estimulados pelo desejo de criar e investigar modelos novos e potencialmente interessantes, e ao mesmo tempo, pela crescente presença de dados policotômicos em avaliações educacionais [van der Linden & Hambleton(1997)].

Andrich (1978) propôs o **Modelo de Escala Gradual**, que pode ser considerado como um caso particular do modelo de Samejima (1969), na medida em que os escores das categorias dos itens são equidistantes (ou seja, igualmente espaçados). Posteriormente, Masters (1982) desenvolveu uma extensão do **modelo de Rasch** (1960), denominado de **Modelo de Crédito Parcial**, o qual, tal como o modelo de Rasch, possui apenas parâmetros relativos a dificuldade (neste caso da categoria, e não do item).

O modelo sugerido por Muraki (1992), conhecido por **Modelo de Crédito Parcial Generalizado**, foi baseado no modelo de Masters (1982), na medida em que introduziu um novo conjunto de parâmetros relativos ao poder de discriminação das categorias.

Um ponto a ser ressaltado, é a pouca utilização destes modelos, muito embora estes apresentem um potencial muito grande no sentido da quantidade de testes que podem ser melhor analisados através dos mesmos.

Neste trabalho focaremos nossa atenção no Modelo de Resposta Nominal (MRN) e no Modelo de Resposta Gradual (MRG). O primeiro por se constituir numa alternativa para análise de itens de múltipla escolha e o segundo por ser aplicável a uma outra classe de itens (por exemplo itens dissertativos). Neste Capítulo e no próximo abordaremos a estimação no Modelo de Resposta Nominal. Já nos outros dois capítulos subseqüentes trataremos do Modelo de Resposta Gradual.

## 4.2 Apresentação do modelo

Sejam  $Y_{ijs}$ , as variáveis que assumem valor 1, se o indivíduo  $j$ ,  $j = 1, 2, \dots, n$  escolhe a categoria  $s$ ,  $s = 1, 2, \dots, m_i$  do item  $i$ ,  $i = 1, 2, \dots, I$  e 0, caso contrário. Sejam ainda  $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijm_i})'$ ,  $\mathbf{Y}_{.j} = (\mathbf{Y}'_{1j}, \dots, \mathbf{Y}'_{ij})'$  e  $\mathbf{Y}_{\dots} = (\mathbf{Y}'_{.1}, \dots, \mathbf{Y}'_{.n})'$ , os vetores de respostas do indivíduo  $j$  ao item  $i$ , o conjunto integral de respostas do indivíduo  $j$  e o conjunto das respostas de todos os indivíduos, respectivamente. Além disso, considere que as categorias de cada um dos itens são **mutuamente exclusivas e exaustivas**.

Os modelos policotômicos possuem, em geral, 3 suposições que são imprescindíveis para a construção do processo de estimação. Além das suposições (S1) e (S2) descritas na Seção 2.1 considera-se uma terceira :

**(S3)** A probabilidade de escolha da categoria  $s$  do item  $i$  pelo indivíduo  $j$ , pode ser representada pelo modelo **multivariado de Bernoulli**

$$P(\mathbf{Y}_{ij} = \mathbf{y}_{ij} | \theta, \zeta_i) \equiv P(\mathbf{Y}_{ij} = \mathbf{y}_{ij}) = \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}},$$

onde  $P_{ijs}$  é uma **função de resposta ao item** relacionada a algum modelo policotômico.

A formulação do MRN é dada por

$$P_{ijs} = P(Y_{ijs} = 1 | \theta_j, \zeta_i) = \frac{e^{a_{is}(\theta_j - b_{is})}}{\sum_{h=1}^{m_i} e^{a_{ih}(\theta_j - b_{ih})}}, \quad (4.1)$$

com  $a_{is}$  e  $b_{is}$  representando os parâmetros relacionados a **discriminação e dificuldade da  $s$  - ésima categoria do  $i$  - ésimo item**. No presente modelo podemos ter valores negativos para ambos os tipos de parâmetros. No caso do parâmetro de discriminação, valores negativos são esperados para as alternativas incorretas e positivos para corretas. Isto significa que quanto maior a habilidade do indivíduo maior a chance do mesmo escolher a alternativa correta (valores



positivos). Com relação ao parâmetro de dificuldade, espera-se que a alternativa correta sempre apresente o maior valor. Isto nos diz que essa alternativa é mais difícil de ser escolhida. Para cada  $\theta_j$  fixado, temos que a soma da probabilidade de escolha entre as  $m_i$  categorias é igual a 1 ( $\sum_{s=1}^{m_i} P_{ijs} = 1$ ). Notemos também que, em um item com  $m_i$  categorias, teremos  $2m_i$  parâmetros a serem estimados. A Figura 4.1 ilustra o comportamento da função de resposta ao item para o MRN de quatro itens (lembrando que estamos considerando a alternativa com os maiores valores dos parâmetros como a alternativa correta). A Figura 4.1(a) apresenta um item relativamente fácil ( $b = 0,3$ , categoria correta). A Figura 4.1(b) apresentam o mesmo item apenas trocando as alternativas corretas (neste item a categoria correta é a C3). A Figura 4.1(c) apresenta o mesmo item com a categoria correta apresentando maior dificuldade ( $b = 1,2$ ) o que faz que a curva dessa categoria se desloque para a direita, exigindo do indivíduo um conhecimento maior para que ele tenha uma chance razoável de escolher essa categoria. A Figura 4.1(d) apresenta o mesmo item com uma das categorias incorretas (C3) com o parâmetro de discriminação um pouco maior ( $a = 1,4$ ) o que faz com que esta categoria fique mais íngreme e também muda as probabilidades de escolha das outras categorias devido a forma da função de resposta ao item do MRN (equação 4.1).

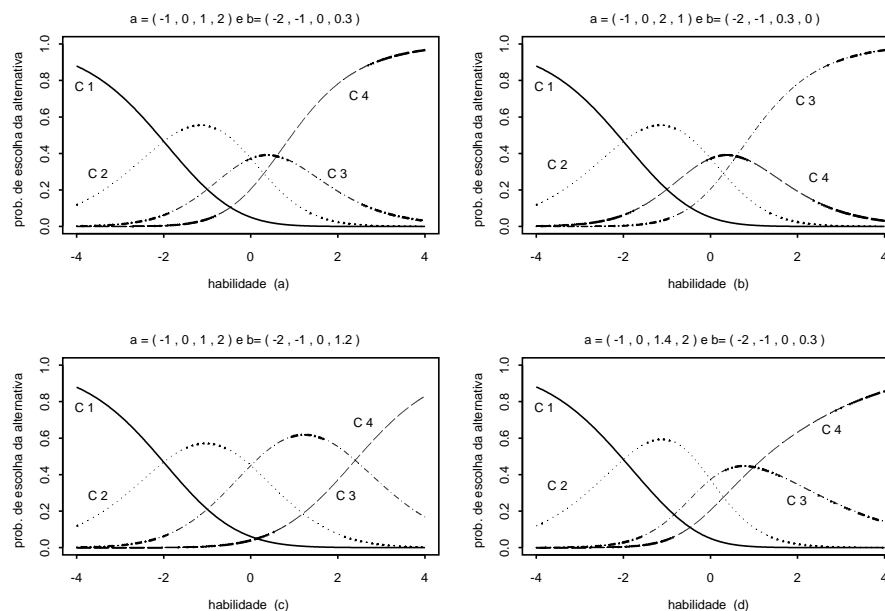


Figura 4.1 Representação gráfica do MRN de quatro itens

Na próxima Seção apresentaremos a estimação dos parâmetros dos itens.

### 4.3 Estimação dos parâmetros dos itens

Tal como nos capítulos anteriores, consideraremos o problema de estimação dos parâmetros dos itens conhecendo e não conhecendo as habilidades dos respondentes. Os dados omissos (falta de resposta de alguns indivíduos) ou respostas nulas (indivíduos que escolhem mais de uma alternativa) serão considerados como uma outra categoria.

#### 4.3.1 Habilidades conhecidas

Como descrito em Baker (1992), devido ao termo  $z_{ijs} = a_{is}(\theta_j - b_{is})$ , aparecer tanto no numerador quanto no denominador (equação (4.1)), existem problemas de indeterminação além daqueles descritos para os modelos dicotômicos [Bock (1972)].

Bock (1972) sugere que devemos impor restrições lineares arbitrárias tais como  $\sum_{s=1}^{m_i} z_{ijs} = 0$ . Para tanto, consideremos primeiramente os seguintes vetores paramétricos :  $\zeta_i = (\mathbf{a}'_i, \mathbf{b}'_i)'$ ,  $\zeta_i^* = (\mathbf{a}'_i, \mathbf{d}'_i)'$  e  $\mathbf{\Gamma}_i = (\boldsymbol{\alpha}'_i, \boldsymbol{\delta}'_i)'$ , que denotam os vetores de parâmetros originais, transformados e irrestritos, respectivamente [Bock (1972)], com  $\mathbf{a}_i = (a_{i1}, \dots, a_{im_i})'$ ,  $\mathbf{b}_i = (b_{i1}, \dots, b_{im_i})'$ ,  $\mathbf{d}_i = (d_{i1}, \dots, d_{im_i})'$ ,  $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{i(m_i-1)})'$ ,  $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{i(m_i-1)})'$ ,  $d_{is} = -a_{is}b_{is}$ ,  $\alpha_{ih} = a_{i1} - a_{i(h+1)}$  e  $\delta_{ih} = d_{i1} - d_{i(h+1)}$ ,  $i = 1, \dots, I$ ,  $s = 1, \dots, m_i$  e  $h = 1, \dots, m_i - 1$ .

Voltando as restrições sugeridas anteriormente, temos que  $\sum_{s=1}^{m_i} z_{ijs} = 0$  implica que

$$\sum_{s=1}^{m_i} a_{is} = 0 \quad e \quad \sum_{s=1}^{m_i} d_{is} = 0. \quad (4.2)$$

Contudo, Bock (1972) afirma que usar uma transformação nos parâmetros  $\zeta^*$ , de modo a contemplar a restrição descrita em (4.2) e desenvolver o processo iterativo usando o novo vetor de parâmetros  $\mathbf{\Gamma}$ , é computacionalmente melhor do que, por exemplo, usar técnicas como a dos **Multiplicadores de Lagrange** usando-se o vetor  $\zeta^*$ .

Dessa forma, descreveremos o processo de implementação de tais restrições, e subseqüentemente desenvolveremos o processo de estimação, propriamente. Para tanto, consideremos:

$$\mathbf{Z}_i = \mathbf{B}\zeta_i^* \mathbf{A}_i, \quad (4.3)$$

com  $\mathbf{B} = (1, \theta)'$ ,  $\mathbf{A}_i = \mathbf{S}_i \mathbf{T}_i$  e

$$\{\mathbf{A}_i\}_{rs} = \Delta_{rs} - \frac{1}{m_i}, \quad (4.4)$$

onde,  $\Delta_{rs} = 1$ , se  $r = s$  e 0, caso contrário (é conhecido como **Delta de Kronecker**) [Bock (1997)]. A matriz  $\mathbf{A}_i$  é não-inversível ( $posto(\mathbf{A}_i) = m_i - 1$ ), o que impossibilita a definição dos

parâmetros irrestritos através da mesma. Uma reparametrização de (4.3) pode ser considerada fazendo-se

$$\mathbf{Z}_i = \mathbf{B}\zeta_i^* \mathbf{S}_i \mathbf{T}_i, \quad (4.5)$$

com

$$\mathbf{S}_i = \begin{bmatrix} 1 & 1 & \dots & 1 \\ -1 & 0 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{bmatrix}$$

e

$$\mathbf{T}_i = (\mathbf{S}'_i \mathbf{S}_i)^{-1} \mathbf{S}'_i \mathbf{A}_i = \begin{bmatrix} \frac{1}{m_i} & \frac{1}{m_i} - 1 & \dots & \dots & \frac{1}{m_i} \\ \frac{1}{m_i} & \dots & \frac{1}{m_i} - 1 & \dots & \frac{1}{m_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m_i} & \frac{1}{m_i} & \dots & \dots & \frac{1}{m_i} - 1 \end{bmatrix}, \quad (4.6)$$

sendo  $\mathbf{\Gamma}_i = \zeta_i^* \mathbf{S}_i$ , os **parâmetros irrestritos**. A matriz  $\mathbf{S}_i$  representa um conjunto de contrastes simples (em nosso caso da categoria 1 contra as demais). Dessa forma, usando as suposições (S1), (S2) e (S3), temos que a verossimilhança é dada por

$$L(\mathbf{\Gamma}) = \prod_{i=1}^I \prod_{j=1}^n \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}, \quad (4.7)$$

e, a log-verossimilhança por

$$l(\mathbf{\Gamma}) = \sum_{i=1}^I \sum_{j=1}^n \sum_{s=1}^{m_i} \{y_{ijs} \ln P_{ijs}\}, \quad (4.8)$$

onde  $P_{ijs} = \frac{e^{z_{ijs}}}{w_{ij}}$  e  $w_{ij} = \sum_{s'=1}^{m_i} e^{z_{ijs'}}$ . Usando (4.5), temos que

$$z_{ijs} = \begin{cases} \sum_{s'=1}^{m_i-1} \frac{\delta_{is'}}{m_i} + \left( \sum_{s'=1}^{m_i-1} \frac{\alpha_{is'}}{m_i} \right) \theta_j, & \text{se } s = 1 \\ \left( -\delta_{i(s-1)} + \sum_{s'=1}^{m_i-1} \frac{\delta_{is'}}{m_i} \right) + \left( \sum_{s'=1}^{m_i-1} \frac{\alpha_{is'}}{m_i} \right) \theta_j, & \text{para } s = 2, \dots, m_i \end{cases} \quad (4.9)$$

Derivando (4.9) com respeito a  $\gamma_{ih} = (\delta_{ih}, \alpha_{ih})'$ , temos (ver desenvolvimento no Apêndice A.2.1)

$$\frac{\partial z_{ijs}}{\partial \delta_{ih}} = \begin{cases} \frac{1}{m_i} - 1 & , \text{ se } s = h + 1 \\ \frac{1}{m_i} & , \text{ se } s \neq h + 1 \end{cases} \quad (4.10)$$

e

$$\frac{\partial z_{ijs}}{\partial \alpha_{ih}} = \begin{cases} \left(\frac{1}{m_i} - 1\right) \theta_j & , \text{ se } s = h + 1 \\ \left(\frac{1}{m_i}\right) \theta_j & , \text{ se } s \neq h + 1 \end{cases} \quad (4.11)$$

Temos também que

$$\frac{\partial w_{ij}}{\partial \delta_{ih}} = \frac{w_{ij}}{m_i} - e^{z_{ij}(h+1)} \quad (4.12)$$

$$\frac{\partial w_{ij}}{\partial \alpha_{ih}} = \left(\frac{w_{ij}}{m_i} - e^{z_{ij}(h+1)}\right) \theta_j \quad (4.13)$$

De forma análoga, temos que

$$\frac{\partial P_{ijs}}{\partial \delta_{ih}} = \begin{cases} -P_{ij(h+1)} (1 - P_{ij(h+1)}) & , \text{ se } s = h + 1 \\ P_{ijs} P_{ij(h+1)} & , \text{ se } s \neq h + 1 \end{cases} \quad (4.14)$$

e

$$\frac{\partial P_{ijs}}{\partial \alpha_{ih}} = \begin{cases} \{-P_{ij(h+1)} (1 - P_{ij(h+1)})\} \theta_j & , \text{ se } s = h + 1 \\ \{P_{ijs} P_{ij(h+1)}\} \theta_j & , \text{ se } s \neq h + 1 \end{cases} \quad (4.15)$$

Assim, usando (4.15), temos para os parâmetros de um determinado item e categoria, que a função escore é dada por :

$$\mathbf{S}(\gamma_{ih}) = (S(\delta_{ih}), S(\alpha_{ih}))', \quad (4.16)$$

com

$$S(\delta_{ih}) = \frac{\partial l(\mathbf{\Gamma})}{\partial \delta_{ih}} = \sum_{j=1}^n \{P_{ij(h+1)} - y_{ij(h+1)}\} \quad (4.17)$$

e

$$S(\alpha_{ih}) = \frac{\partial l(\boldsymbol{\Gamma})}{\partial \alpha_{ih}} = \sum_{j=1}^n \theta_j \{P_{ij(h+1)} - y_{ij(h+1)}\}. \quad (4.18)$$

Seguindo Baker(1992), temos que (para um determinado item), o vetor escore é dado por

$$\mathbf{S}(\boldsymbol{\Gamma}_i) = (\mathbf{S}(\boldsymbol{\delta}_i)', \mathbf{S}(\boldsymbol{\alpha}_i)')', \quad (4.19)$$

com

$$\mathbf{S}(\boldsymbol{\delta}_i) = \sum_{j=1}^n \mathbf{T}_i \{\mathbf{y}_{ij.} - \mathbf{P}_{ij.}\} \quad (4.20)$$

e

$$\mathbf{S}(\boldsymbol{\alpha}_i) = \sum_{j=1}^n \mathbf{T}_i \{\mathbf{y}_{ij.} - \mathbf{P}_{ij.}\} \theta_j, \quad (4.21)$$

o que implica que

$$\mathbf{S}(\boldsymbol{\Gamma}_i) = \sum_{j=1}^n \left\{ \begin{bmatrix} 1 \\ \theta_j \end{bmatrix} \otimes \mathbf{T}_i [\mathbf{y}_{ij.} - \mathbf{P}_{ij.}] \right\}. \quad (4.22)$$

com  $\mathbf{y}_{ij.} = (y_{ij1}, \dots, y_{ijm_i})'$ ,  $\mathbf{P}_{ij.} = (P_{ij1}, \dots, P_{ijm_i})'$ , e  $\otimes$  denotando o **produto de Kronecker** à esquerda [Harville (1997)].

Como o sistema de equações obtido a partir de (4.22) não possui solução explícita, temos que usar algum método iterativo. A matriz Hessiana é dada por (ver desenvolvimento no Apêndice A.2.1)

$$\mathbf{H}(\boldsymbol{\Gamma}_i) = - \sum_{j=1}^n \left\{ \begin{bmatrix} 1 & \theta_j \\ \theta_j & \theta_j^2 \end{bmatrix} \otimes [\mathbf{T}_i \mathbf{W}_{ij} \mathbf{T}_i'] \right\}, \quad (4.23)$$

com  $\{\mathbf{W}_{ij}\}_{rs} = P_{ijr}(\Delta_{rs} - P_{ijs})$ , onde  $\Delta_{rs}$  é como em (4.4). Como podemos notar, a matriz Hessiana descrita em (4.23) é não estocástica e, portanto a Informação de Fisher nada mais é que  $\mathbf{I}(\boldsymbol{\Gamma}_i) = -\mathbf{H}(\boldsymbol{\Gamma}_i)$ . Isso conduz a equivalência de ambos os processos iterativos (Newton-Raphson/Escore de Fisher).

Dessa forma, considerando  $\hat{\boldsymbol{\Gamma}}_i^{(t)}$ , uma estimativa de  $\boldsymbol{\Gamma}$  na iteração  $t$ , temos que o algoritmo Newton-Raphson/Escore de Fisher é dado por

$$\hat{\Gamma}_i^{(t+1)} = \hat{\Gamma}_i^{(t)} - \left[ \mathbf{H} \left( \hat{\Gamma}_i^{(t)} \right) \right]^{-1} \left[ \mathbf{S} \left( \hat{\Gamma}_i^{(t)} \right) \right]$$

$t = 0, 1, 2, \dots$

Detalhes dos valores iniciais para os processos iterativos podem ser encontrados em Muraki & Bock (1997). Basicamente são consideradas extensões dos valores usados nos modelos dicotômicos.

### Habilidades conhecidas e agrupadas

Considerando a abordagem de agrupamento das habilidades descrito na Seção 2.2.1, temos que a verossimilhança pode ser escrita como

$$L(\Gamma) = \prod_{i=1}^I \prod_{l=1}^q \prod_{s=1}^{m_i} \left\{ \frac{f_{il}}{\prod_{h=1}^{m_i} r_{ilh}} \right\} \{P_{ils}^{r_{ils}}\}, \quad (4.24)$$

e a log-verossimilhança como

$$l(\Gamma) = const + \sum_{i=1}^I \sum_{l=1}^q \sum_{s=1}^{m_i} \{r_{ils} \ln P_{ils}\}, \quad (4.25)$$

com  $f_{ils}$  e  $r_{ils}$ , sendo o número de indivíduos com nível de habilidade em torno de  $\bar{\theta}_l$  que respondem ao item  $i$ , e dentro desses indivíduos, o número daqueles que escolhem a categoria  $s$ , respectivamente.

Dessa forma, diferenciando-se (4.25) com relação aos parâmetros dos itens, obtêm-se as funções de estimação (vetor escore)

$$\mathbf{S}(\Gamma_i) = \sum_{l=1}^q \sum_{s=1}^{m_i} \left\{ \frac{r_{ils}}{P_{ils}} \left( \frac{\partial P_{ils}}{\partial \Gamma_i} \right) \right\}. \quad (4.26)$$

Usando desenvolvimento análogo ao da abordagem de não agrupamento das habilidades, temos que o vetor escore é dado por

$$\mathbf{S}(\Gamma_i) = \sum_{l=1}^q \left\{ \left[ \begin{array}{c} 1 \\ \bar{\theta}_l \end{array} \right] \otimes \mathbf{T}_i [\mathbf{r}_{il.} - f_{il} \mathbf{P}_{il.}] \right\}, \quad (4.27)$$

com  $\mathbf{r}_{il.} = (r_{il1}, \dots, r_{ilm_i})'$  e  $\mathbf{P}_{il.} = (P_{il1}, \dots, P_{ilm_i})'$ , em que  $P_{ils}$  é como em (4.1), com  $\theta_i$  substituído por  $\bar{\theta}_l$ .

Segue-se então, que a matriz Hessiana é dada por

$$\mathbf{H}(\boldsymbol{\Gamma}_i) = - \sum_{l=1}^q \left\{ \begin{bmatrix} 1 & \bar{\theta}_l \\ \bar{\theta}_l & \bar{\theta}_l^2 \end{bmatrix} \otimes [\mathbf{T}_i \mathbf{W}_{il} \mathbf{T}'_i] \right\}. \quad (4.28)$$

com  $\{\mathbf{W}_{il}\}_{rs} = f_{il} [P_{ilr} (\Delta_{rs} - P_{ils})]$ , com  $\Delta_{rs}$  como definido em (4.4). Novamente, pelo fato da matriz Hessiana ser não-estocástica, temos que  $\mathbf{I}(\boldsymbol{\Gamma}_i) = -\mathbf{H}(\boldsymbol{\Gamma}_i)$ , e dessa forma, podemos executar usualmente os procedimentos iterativos. Na próxima Seção abordaremos o procedimento de estimação ante o desconhecimento das habilidades.

### 4.3.2 Habilidades desconhecidas

De modo análogo aos modelos anteriores, devemos construir a probabilidade marginal de respostas para um determinado indivíduo de tal forma a eliminar as habilidades, numa primeira etapa, do processo de estimação [Bock (1997)]. Assim sendo, temos que a referida probabilidade é dada por

$$P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \boldsymbol{\Gamma}, \boldsymbol{\eta}) = \int_{\mathbb{R}} P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\Gamma}) g(\theta | \boldsymbol{\eta}) d\theta \quad (4.29)$$

Portanto, usando (4.29), temos que a verossimilhança marginal é dada por

$$L(\boldsymbol{\Gamma}, \boldsymbol{\eta}) = \prod_{j=1}^n P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \boldsymbol{\Gamma}, \boldsymbol{\eta}) \quad (4.30)$$

e a log-verossimilhança por

$$l(\boldsymbol{\Gamma}, \boldsymbol{\eta}) = \sum_{j=1}^n \ln P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \boldsymbol{\Gamma}, \boldsymbol{\eta}) \quad (4.31)$$

Derivando-se (4.31) com relação aos parâmetros dos itens, obtem-se o sistema de equações de estimação, o qual (ver Apêndice A.2.1) é dado por

$$\mathbf{S}(\boldsymbol{\Gamma}_i) = \frac{\partial l(\boldsymbol{\Gamma}, \boldsymbol{\eta})}{\partial \boldsymbol{\Gamma}_i} = \sum_{l=1}^q \left\{ \begin{bmatrix} 1 \\ \bar{\theta}_l \end{bmatrix} \otimes \mathbf{T}_i [\bar{\mathbf{r}}_{il} - \bar{f}_{il} \mathbf{P}_{il}] \right\} \quad (4.32)$$

com  $\mathbf{r}_{il} = (\bar{r}_{il1}, \dots, \bar{r}_{ilm_i})'$  e  $\mathbf{P}_{il} = (P_{il1}, \dots, P_{ilm_i})'$ , com  $P_{ils}$ , como definido em (4.1), com  $\theta_j$  substituído por  $\bar{\theta}_l$ ,  $\bar{r}_{ils} = \sum_{j=1}^n y_{ijs} g_j^*(\bar{\theta}_l)$  e  $g_j^*(\bar{\theta}_l) = \frac{P(\mathbf{Y}_{.j} | \boldsymbol{\Gamma}, \boldsymbol{\eta}) g(\bar{\theta}_l | \boldsymbol{\eta})}{\sum_{l=1}^q P(\mathbf{Y}_{.j} | \boldsymbol{\Gamma}, \bar{\theta}_l) g(\bar{\theta}_l | \boldsymbol{\eta})}$ .

Para resolver o sistema de equações definido em (4.32) Bock(1997) propõe uma adaptação do **algoritmo EM**, à semelhança dos modelos dicotômicos.

#### Adaptação do Algoritmo EM

A proposta de Bock constitui-se no seguinte procedimento

**Passo E**

Usar os pontos de quadratura  $\bar{\theta}_l$ , os pesos  $A_l^{(t)}$ ,  $l = 1, \dots, q$  e estimativas provisórias dos parâmetros dos itens,  $\hat{\zeta}_i^{(t)}$ ,  $i = 1, \dots, I$ , para gerar  $g_j^*(\bar{\theta}_l)^{(t)}$  e, posteriormente,  $\bar{r}_{ils}^{(t)}$  e  $\bar{f}_{il}^{(t)}$ ,  $i = 1, \dots, I$ ,  $l = 1, \dots, q$  e  $s = 0, 1, \dots, m_i$ .

**Passo M**

Com  $\bar{\mathbf{r}}^{(t)}$  e  $\bar{\mathbf{f}}^{(t)}$  obtidos no Passo E, resolver as equações de estimação para  $\Gamma_i$ ,  $i = 1, \dots, I$ , usando o algoritmo de Newton-Raphson ou Escore de Fisher.

Para efetuar os procedimentos iterativos, podemos usar a expressão (4.32) (**vetor escore**) e (4.28) (**matriz Hessiana**) ( $\mathbf{I}(\Gamma_i) = -\mathbf{H}(\Gamma_i)$ ), substituindo-se  $\mathbf{r}$ ,  $\mathbf{f}$  e  $\boldsymbol{\theta}$  por  $\bar{\mathbf{r}}$ ,  $\bar{\mathbf{f}}$  e  $\bar{\boldsymbol{\theta}}$ , respectivamente, ou seja

$$\mathbf{H}(\Gamma_i) = \frac{\partial l(\Gamma, \boldsymbol{\eta})}{\partial \Gamma_i \partial \Gamma_i'} = - \sum_{l=1}^q \left\{ \begin{bmatrix} 1 & \bar{\theta}_l \\ \bar{\theta}_l & \bar{\theta}_l^2 \end{bmatrix} \otimes [\mathbf{T}_i \mathbf{W}_{il} \mathbf{T}_i'] \right\}. \quad (4.33)$$

Tem-se também que a estimação é feita para cada item em separado. Este procedimento gera estimativas consistentes. A **superfície de verossimilhança** é **convexa** em quase toda parte. Dessa forma o vetor de **pontos estacionários** será único e portanto o máximo global da função de verossimilhança [Bock & Zimowsk (1997)], [Bock(1997)]. O processo de estimação convergirá a menos de padrões de resposta aberrantes [Bock & Zimowsk (1997)]. A informação de Fisher fornecerá (desde que esta seja positiva definida) os erros-padrão assintóticos das estimativas dos parâmetros dos itens.

Na próxima Seção apresentaremos a estimação das habilidades.

**4.4 Estimação das habilidades**

Da mesma forma que nos capítulos anteriores, consideraremos a estimação individual das habilidades. Para realizar este processo, consideraremos também que os parâmetros dos itens são conhecidos, e para tanto utilizaremos os parâmetros irrestritos. Para um determinado indivíduo, temos que a verossimilhança é dada por

$$L(\theta_j) = \prod_{i=1}^I \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}, \quad (4.34)$$

o que nos leva a seguinte log-verossimilhança

$$l(\theta_j) = \sum_{i=1}^I \sum_{s=1}^{m_i} y_{ijs} \ln P_{ijs}, \quad (4.35)$$



Por outro lado, utilizando (4.9), temos que

$$\frac{\partial z_{ijs}}{\partial \theta_j} = \begin{cases} \sum_{s'=1}^{m_i} \frac{\alpha_{is'}}{m_i} & , \text{ se } s = 1 \\ -\alpha_{i(s-1)} + \sum_{s'=1}^{m_i-1} \frac{\alpha_{is'}}{m_i} & , \text{ para } s = 2, \dots, m_i \end{cases} \quad (4.36)$$

e

$$\frac{\partial w_{ij}}{\partial \theta_j} = \left\{ \sum_{s'=1}^{m_i} e^{z_{ijs'}} \right\} \left\{ \sum_{s'=1}^{m_i-1} \frac{\alpha_{is'}}{m_i} \right\} - \sum_{s'=1}^{m_i-1} e^{z_{ij(s'+1)}} , \quad (4.37)$$

Além disso, temos que

$$\frac{\partial P_{ijs}}{\partial \theta_j} = \begin{cases} P_{ij1} \left\{ \sum_{s'=1}^{m_i-1} P_{ij(s'+1)} \alpha_{is'} \right\} & , \text{ se } s = 1 \\ P_{ijs} \left\{ \alpha_{i(s-1)} + \sum_{s'=1}^{m_i-1} P_{ij(s'+1)} \alpha_{is'} \right\} & , \text{ para } s = 2, \dots, m_i \end{cases} \quad (4.38)$$

O estimador de máxima verossimilhança das habilidades é fornecido pela solução do sistema de equações de estimação, qual seja :

$$S(\theta_j) = \sum_{i=1}^I \sum_{s=1}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial \theta_j} \right) \right\} . \quad (4.39)$$

Que resulta em (ver desenvolvimentos no Apêndice A.2.3)

$$S(\theta_j) = - \sum_{i=1}^I \sum_{s=1}^{m_i-1} \alpha_{is} (y_{ij(s+1)} - P_{ij(s+1)}) . \quad (4.40)$$

Temos que, em forma matricial, (4.40) pode ser escrita como [Baker (1992)]

$$S(\theta_j) = \sum_{i=1}^I \alpha'_i \mathbf{T}_i [\mathbf{y}_{ij.} - \mathbf{P}_{ij.}] . \quad (4.41)$$

Como vemos, pela não-linearidade do sistema de equações definido em (4.41), devemos usar algum método iterativo para resolvê-lo. Para tanto, temos que a Matriz Hessiana é dada por

$$H(\theta_j) = \sum_{i=1}^I \left\{ \left( \sum_{h=1}^{m_i-1} P_{ij(h+1)} \alpha_{ih} \right)^2 - \sum_{h=1}^{m_i-1} \alpha_{ih}^2 P_{ij(h+1)} \right\},$$

que em forma matricial pode ser escrita como [Baker (1992)]

$$H(\theta_j) = - \sum_{i=1}^I \{ \boldsymbol{\alpha}'_i \mathbf{T}_i \mathbf{W}_{ij} \mathbf{T}'_i \boldsymbol{\alpha}_i \}. \quad (4.42)$$

Dessa a forma os procedimentos iterativos podem ser implementados como em (2.32) e (2.33). A **superfície de verossimilhança** é convexa a menos de um conjunto de medida nula. Isto implica na convergência do processo a menos de padrões de resposta aberrantes (indivíduos com habilidades altas que escolhem alternativas incorretas, por exemplo) [Baker (1997)].

---

## Capítulo 5

# Modelo de Resposta Nominal: Estimação Bayesiana

---

### 5.1 Introdução

Basicamente, o objetivo do desenvolvimento dos métodos bayesianos no contexto dos modelos de respostas policotômicas, caminham no mesmo sentido dos modelos dicotômicos. Dito isso, a abordagem deste capítulo seguirá a mesma de seus antecessores.

### 5.2 Estimação dos parâmetros dos itens

Considerando-se que os parâmetros de diferentes itens são **estocasticamente independentes**, ou seja, não trazem informação um dos outros, atribuiremos prioris independentes. No entanto, como o vetor de parâmetros irrestritos ( $\mathbf{\Gamma}_i$ ) pode assumir qualquer valor no  $\mathbb{R}^{2m_i-1}$ , atribuiremos uma priori multivariada para os parâmetros de cada item. Mais especificamente, consideraremos que  $\mathbf{\Gamma}_i \sim N_{(2m_i-2)}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ .

#### 5.2.1 Habilidades conhecidas

Usando notações semelhantes às do Capítulo 3, temos que, a posteriori pode ser escrita como

$$\begin{aligned} f(\mathbf{\Gamma}|\mathbf{y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau}) &\propto L(\mathbf{\Gamma}) f(\mathbf{\Gamma}|\boldsymbol{\tau}) f(\boldsymbol{\tau}) \\ &= \left\{ \prod_{i=1}^I L(\mathbf{\Gamma}_i) \right\} \left\{ \prod_{i=1}^I f(\mathbf{\Gamma}_i|\boldsymbol{\tau}_i) \right\} \left\{ \prod_{i=1}^I f(\boldsymbol{\tau}_i) \right\}, \end{aligned}$$

e o logaritmo da posteriori como

$$\ln f(\mathbf{\Gamma}|\mathbf{y}_{..}, \boldsymbol{\theta}, \boldsymbol{\tau}) \propto \sum_{i=1}^I \ln L(\mathbf{\Gamma}_i) + \sum_{i=1}^I \ln f(\mathbf{\Gamma}_i|\boldsymbol{\tau}_i) + \sum_{i=1}^I \ln f(\boldsymbol{\tau}_i), \quad (5.1)$$

onde  $\boldsymbol{\tau}$  é o vetor de hiperparâmetros associados a  $\mathbf{\Gamma}$ . Como estimadores dos parâmetros dos itens, usaremos a **moda a posteriori**. Para tal devemos encontrar o máximo de (5.1), que é obtido solucionando-se o seguinte sistema de equações bayesianas

$$S(\boldsymbol{\Gamma}_i)_B = \frac{\partial \ln L(\boldsymbol{\Gamma}_i)}{\partial \boldsymbol{\Gamma}_i} + \frac{\partial \ln f(\boldsymbol{\Gamma}_i | \boldsymbol{\tau}_i)}{\partial \boldsymbol{\Gamma}_i}. \quad (5.2)$$

A primeira parcela de (5.2) gera as funções de estimação definidas em (4.22). Para a segunda parcela, considerando a priori definida anteriormente, temos que

$$f(\boldsymbol{\Gamma}_i | \boldsymbol{\tau}_i) = \frac{1}{2\pi |\boldsymbol{\Sigma}_i|^{\frac{p}{2}}} \exp \left\{ -\frac{1}{2} [(\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i)] \right\} \quad (5.3)$$

e

$$\ln f(\boldsymbol{\Gamma}_i | \boldsymbol{\tau}_i) \propto -\frac{1}{2} [(\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i)]. \quad (5.4)$$

Usando resultados de Mardia et al (1979), temos que

$$\frac{\partial \ln f(\boldsymbol{\Gamma}_i | \boldsymbol{\tau}_i)}{\partial \boldsymbol{\Gamma}_i} = -\mathbf{V}_i (\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i) \quad (5.5)$$

e

$$\frac{\partial \ln f(\boldsymbol{\Gamma}_i | \boldsymbol{\tau}_i)}{\partial \boldsymbol{\Gamma}_i \partial \boldsymbol{\Gamma}_i'} = -\mathbf{V}_i, \quad (5.6)$$

com  $\mathbf{V}_i = \boldsymbol{\Sigma}_i^{-1}$ . Dessa forma, usando (4.22), (4.23), (5.5) e (5.6), temos que o vetor escore e a matriz Hessiana são dados, respectivamente, por

$$\mathbf{S}(\boldsymbol{\Gamma}_i)_B = \sum_{j=1}^n \left\{ \begin{bmatrix} 1 \\ \theta_j \end{bmatrix} \otimes \mathbf{T}_i [\mathbf{y}_{ij} - \mathbf{P}_{ij}] \right\} - \mathbf{V}_i (\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i) \quad (5.7)$$

e

$$\mathbf{H}(\boldsymbol{\Gamma}_i)_B = -\sum_{j=1}^n \left\{ \begin{bmatrix} 1 & \theta_j \\ \theta_j & \theta_j^2 \end{bmatrix} \otimes [\mathbf{T}_i \mathbf{W}_{ij} \mathbf{T}_i'] \right\} - \mathbf{V}_i. \quad (5.8)$$

Como no caso da estimação por MV (Subseção 4.3.2), a Informação de Fisher é dada pelo simétrico da matriz Hessiana (devido a mesma ser não-estocástica). Dessa forma, os métodos iterativos podem ser implementados de maneira usual.

### Habilidades conhecidas e agrupadas

Usando (4.27), (4.28), (5.5) e (5.6), temos que as expressões para o processo de estimação são dadas por

$$\mathbf{S}(\Gamma_i) = \sum_{l=1}^q \left\{ \left[ \begin{array}{c} 1 \\ \bar{\theta}_l \end{array} \right] \otimes \mathbf{T}_i [\mathbf{r}_{il} - f_{il} \mathbf{P}_{il}] \right\} - \mathbf{V}_i(\Gamma_i - \boldsymbol{\mu}_i)$$

e

$$\mathbf{H}(\Gamma_i) = - \sum_{l=1}^q \left\{ \left[ \begin{array}{cc} 1 & \bar{\theta}_l \\ \bar{\theta}_l & \bar{\theta}_l^2 \end{array} \right] \otimes [\mathbf{T}_i \mathbf{W}_{il} \mathbf{T}_i'] \right\} - \mathbf{V}_i.$$

Na próxima Seção abordaremos a estimação dos parâmetros dos itens ante o desconhecimento das habilidades.

### 5.2.2 Habilidades desconhecidas

De conformidade com os modelos dicotômicos, usaremos o método de **Estimação Bayesiana Marginal**. Dessa forma, usando notações semelhantes ao Capítulo 3, temos que a distribuição a posteriori pode ser escrita como

$$f(\Gamma, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{y}_{\dots}) \propto L(\mathbf{y}_{\dots}; \Gamma, \boldsymbol{\theta}) f(\Gamma | \boldsymbol{\tau}) f(\boldsymbol{\tau}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) g(\boldsymbol{\eta}). \quad (5.9)$$

De modo análogo ao Capítulo 3, consideraremos o vetor  $\boldsymbol{\psi} = (\boldsymbol{\Gamma}', \boldsymbol{\eta}')'$  como parâmetros de interesse (muito embora este último não precise ser estimado), e o vetor  $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\tau}')'$  como parâmetros de perturbação. Portanto, integrando (5.9) com relação a  $\boldsymbol{\phi}$ , temos que

$$\begin{aligned} f(\boldsymbol{\psi} | \mathbf{y}_{\dots}) &\propto g(\boldsymbol{\eta}) \left\{ \int L(\mathbf{y}_{\dots}; \Gamma, \boldsymbol{\theta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta} \right\} \left\{ \int f(\Gamma | \boldsymbol{\tau}) f(\boldsymbol{\tau}) d\boldsymbol{\tau} \right\} \\ &= g(\boldsymbol{\eta}) L(\Gamma, \boldsymbol{\eta}) f(\Gamma), \end{aligned} \quad (5.10)$$

com  $L(\Gamma, \boldsymbol{\eta}) = \int L(\mathbf{y}_{\dots}; \Gamma, \boldsymbol{\theta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}$  e  $f(\Gamma) = \int f(\Gamma | \boldsymbol{\tau}) f(\boldsymbol{\tau}) d\boldsymbol{\tau}$ . Notemos que o processo de integração descrito anteriormente não elimina a necessidade de especificação dos **hiperparâmetros** [Baker(1992)].

Com o intuito de calcular a moda da distribuição a posteriori, tomemos convenientemente seu logaritmo natural

$$\ln f(\boldsymbol{\psi} | \mathbf{y}_{\dots}) \propto l(\Gamma, \boldsymbol{\eta}) + \ln f(\Gamma), \quad (5.11)$$

o que gera o seguinte sistema de equações bayesianas

$$\mathbf{S}(\Gamma_i)_{BM} = \frac{\partial l(\Gamma, \boldsymbol{\eta})}{\partial \Gamma_i} + \frac{\partial \ln f(\Gamma)}{\partial \Gamma_i}. \quad (5.12)$$

Podemos notar que a primeira parcela de (5.12) é como definida em (4.32) e que a segunda é como em (5.5). Dessa forma, (5.12), em forma de pontos de quadratura, torna-se

$$\mathbf{S}(\boldsymbol{\Gamma}_i)_{BM} = \sum_{l=1}^q \left\{ \left[ \frac{1}{\bar{\theta}_l} \right] \otimes \mathbf{T}_i [\bar{\mathbf{r}}_{il} - \bar{f}_{il} \mathbf{P}_{il}] \right\} - \mathbf{V}_i (\boldsymbol{\Gamma}_i - \boldsymbol{\mu}_i). \quad (5.13)$$

Pela não linearidade do sistema de equações definido em (5.13), devemos usar algum método iterativo. Dessa forma, usando (4.33) e (5.6), temos que a matriz Hessiana é dada por

$$\mathbf{H}(\boldsymbol{\Gamma}_i)_{BM} = \frac{\partial^2 l(\boldsymbol{\Gamma}, \boldsymbol{\eta})}{\partial \boldsymbol{\Gamma}_i \partial \boldsymbol{\Gamma}_i'} = - \sum_{l=1}^q \left\{ \left[ \frac{1}{\bar{\theta}_l} \quad \bar{\theta}_l \right] \otimes [\mathbf{T}_i \mathbf{W}_{il} \mathbf{T}_i'] \right\} - \mathbf{V}_i, \quad (5.14)$$

e a Informação de Fisher, também neste caso, é dada por  $\mathbf{I}(\boldsymbol{\Gamma}_i) = -\mathbf{H}(\boldsymbol{\Gamma}_i)$ .

O **algoritmo EM** pode ser implementado tal como descrito na Seção 4.3.2, usando a esperança da log-posteriori, ao invés da log-verossimilhança. Os comentários feitos naquela situação também se aplicam na **estimação bayesiana**, porém as estimativas sempre podem ser obtidas independentemente do padrão de respostas. Na próxima Seção apresentaremos o processo de estimação das habilidades.

### 5.3 Estimação das habilidades

O procedimento de estimação bayesiana das habilidades, nos **modelos de respostas policotômicas**, é análogo ao procedimento nos **modelos de respostas dicotômicas**. Como estimadores também adotaremos a moda e a média a posteriori. Sendo assim, temos que a distribuição a posteriori para o traço latente de um determinado indivíduo é dada por

$$g_j^*(\theta_j) \equiv C g(\theta_j | \mathbf{y}_{.j}, \boldsymbol{\Gamma}, \boldsymbol{\eta}) = C P(\mathbf{Y}_{.j} | \theta_j, \boldsymbol{\Gamma}) g(\theta_j | \boldsymbol{\eta}) \quad (5.15)$$

$$\propto P(\mathbf{Y}_{.j} | \theta_j, \boldsymbol{\Gamma}) g(\theta_j | \boldsymbol{\eta}). \quad (5.16)$$

Tal como nos modelos dicotômicos, consideraremos a estimação pela média e moda da distribuição a posteriori.

#### Estimação pela moda a posteriori - MAP

Tomando o logaritmo de (5.15), vem que

$$\ln g_j^*(\theta_j) \equiv l_j^*(\theta_j) \propto l(\theta_j) + \ln g(\theta_j | \boldsymbol{\eta}). \quad (5.17)$$

Segue, de (5.17), que a **equação de estimação bayesiana**, para uma determinada habilidade, é dada por

$$S(\theta_j)_B = \frac{\partial l_j^*(\theta_j)}{\partial \theta_j} = \frac{\partial l(\theta_j)}{\partial \theta_j} + \frac{\partial \ln g(\theta_j|\boldsymbol{\eta})}{\partial \theta_j}. \quad (5.18)$$

Notando que o primeiro termo de (5.18) é dado por (4.41), e adotando uma priori  $N(\mu_\theta, \sigma_\theta^2)$ , temos que (5.18) pode ser escrita como

$$S(\theta_j)_B = \sum_{i=1}^I \boldsymbol{\alpha}'_i \mathbf{T}_i [\mathbf{y}_{ij.} - \mathbf{P}_{ij.}] - \frac{\theta_j - \mu_\theta}{\sigma_\theta^2}. \quad (5.19)$$

Como o sistema definido em (5.19) não possui solução explícita, usaremos algum método iterativo. De (4.42) e (3.32), temos que a matriz Hessiana é dada por

$$H(\theta_j)_B = - \sum_{i=1}^I \{ \boldsymbol{\alpha}'_i \mathbf{T}_i \mathbf{W}_{ij} \mathbf{T}'_i \boldsymbol{\alpha}_i \} - \frac{1}{\sigma_\theta^2}. \quad (5.20)$$

A Informação de Fisher, neste caso, é igual a  $I(\theta_j) = -H(\theta_j)$  e, dessa forma, os procedimentos iterativos podem ser operacionalizados como em (2.32) e (2.33).

#### Estimação pela média a posteriori - EAP

De (5.15), temos que a esperança a posteriori é dada por

$$\mathbb{E}[\theta_j | \mathbf{y}_{.j.}, \boldsymbol{\Gamma}, \boldsymbol{\eta}] = \frac{\int_{\mathbb{R}} \theta P(\mathbf{Y}_{.j.} | \theta, \boldsymbol{\Gamma}) g(\theta | \boldsymbol{\eta}) d\theta}{\int_{\mathbb{R}} P(\mathbf{Y}_{.j.} | \theta, \boldsymbol{\Gamma}) g(\theta | \boldsymbol{\eta}) d\theta} \quad (5.21)$$

Podemos notar que as integrais em (5.21) não possuem solução explícita, portanto devemos usar algum procedimento numérico. Utilizando o método da **Quadratura Gaussiana** descrito na Seção 2.2.2, temos que (5.21) se transforma em

$$\begin{aligned} \mathbb{E}[\theta_j | \mathbf{y}_{.j.}, \boldsymbol{\Gamma}, \boldsymbol{\eta}] &\approx \mathbb{E}[\bar{\theta}_j | \mathbf{y}_{.j.}, \boldsymbol{\Gamma}, \boldsymbol{\eta}] = \frac{\sum_{l=1}^q \bar{\theta}_l P(\mathbf{Y}_{.j.} | \bar{\theta}_l, \boldsymbol{\Gamma}) g(\bar{\theta}_l | \boldsymbol{\eta})}{\sum_{l=1}^q P(\mathbf{Y}_{.j.} | \bar{\theta}_l, \boldsymbol{\Gamma}) g(\bar{\theta}_l | \boldsymbol{\eta})} \\ &= \frac{\sum_{l=1}^q \bar{\theta}_l P(\mathbf{Y}_{.j.} | \bar{\theta}_l, \boldsymbol{\Gamma}) A_l}{\sum_{l=1}^q P(\mathbf{Y}_{.j.} | \bar{\theta}_l, \boldsymbol{\Gamma}) A_l} \end{aligned} \quad (5.22)$$

A expressão (5.22) tem algumas características interessantes. Além de ter a vantagem de ser calculada diretamente, ou seja, sem a necessidade de métodos iterativos, as quantidades necessárias para seu cálculo são um produto final da etapa de estimação dos parâmetros dos itens. Por isso, Mislevy & Stocking (1989), recomendam seu uso para a estimação das habilidades, tal como nos modelos de respostas dicotômicas.

Um medida de precisão do EAP pode ser obtida via variância a posteriori

$$Var [\theta_j | \mathbf{y}_{.j}, \mathbf{\Gamma}, \boldsymbol{\eta}] = \frac{\int_{\mathcal{R}} [\theta - \mathbb{E}(\theta)]^2 P(\mathbf{Y}_{.j} | \theta, \mathbf{\Gamma}) g(\theta | \boldsymbol{\eta}) d\theta}{\int_{\mathcal{R}} P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\zeta}) g(\theta | \boldsymbol{\eta}) d\theta}$$

Em termos dos pontos de quadratura, temos que

$$Var [\bar{\theta}_j | \mathbf{y}_{.j}, \mathbf{\Gamma}, \boldsymbol{\eta}] = \frac{\sum_{l=1}^q \{\bar{\theta}_l - \mathbb{E}[\bar{\theta}_l | \mathbf{y}_{.j}, \boldsymbol{\zeta}, \boldsymbol{\eta}]\}^2 P(\mathbf{Y}_{.j} | \bar{\theta}_l, \mathbf{\Gamma}) g(\bar{\theta}_l | \boldsymbol{\eta})}{\sum_{l=1}^q P(\mathbf{Y}_{.j} | \bar{\theta}_l, \mathbf{\Gamma}) g(\bar{\theta}_l | \boldsymbol{\eta})} \quad (5.23)$$

Para a escolha dos pesos  $A_l$ , podemos usar **pesos teóricos** (assumindo uma distribuição teórica para as habilidades), **pesos empíricos** (estimados durante a calibração dos itens) [Mislevy (1984)] ou **pesos subjetivos** (fornecidos pelo pesquisador) [Mislevy & Bock (1990)].



---

## Capítulo 6

# Modelo de Resposta Gradual: Apresentação e Estimação por Máxima Verossimilhança

---

### 6.1 Introdução

Neste capítulo introduziremos o modelo de Resposta Gradual (MRG) e trataremos da estimação por máxima verossimilhança de seus parâmetros.

### 6.2 Apresentação do modelo

Suponha que as categorias de um determinado item  $i$  possam ser arranjadas em ordem crescente e denotados por  $s_i = 0, 1, \dots, m_i$ , com  $m_i + 1$  sendo o número de categorias do  $i$ -ésimo item. A probabilidade de um indivíduo  $j$  **escolher uma categoria  $s$  ou outra mais alta (maior)** é dado por (uma extensão do **modelo logístico de 2 parâmetros**)

$$P_{i,s}^+(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{is})}}$$

com  $i = 1, \dots, I$ ,  $j = 1, \dots, n$  e  $s = 1, \dots, m_i$ ,  $a_i$  o parâmetro associado a discriminação do item,  $b_{is}$  **o parâmetro associado à dificuldade da categoria  $s$**  e  $D$  e  $\theta_j$  como em (2.1). No caso deste modelo, cabe dizer que a discriminação de uma categoria específica depende tanto do parâmetro de inclinação ( $a_i$ ), comum a todas as categorias do item, quanto da distância das categorias de dificuldade adjacentes [Andrade, Tavares & Valle (2000)]. Pela definição de ordem entre as categorias, devemos ter

$$b_{i1} \leq \dots \leq b_{im_i}$$

ou seja, deve-se ter uma **ordenação** entre os níveis de dificuldade das categorias de um determinado item.

A probabilidade de um indivíduo  $j$  responder a categoria  $s$  no item  $i$  é dada por

$$P_{ijs} = P_{i,s}(\theta_j) = P_{i,s}^+(\theta_j) - P_{i,s+1}^+(\theta_j)$$

Define-se também que

$$P_{i,0}^+(\theta_j) = 1$$

e

$$P_{i,m_i+1}^+(\theta_j) = 0$$

Dessa forma,

$$P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j) = 1 - P_{i,1}^+(\theta_j)$$

e

$$P_{i,m_i}(\theta_j) = P_{i,m_i}^+(\theta_j) - P_{i,m_i+1}^+(\theta_j) = P_{i,m_i}^+(\theta_j)$$

fazendo com que

$$P_{ijs} = \frac{1}{1 + e^{-Da_i(\theta_j - b_{is})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i(s+1)})}} \quad (6.1)$$

Notemos que, para um item com  $m_i + 1$  categorias, teremos  $m_i + 1$  parâmetros para serem estimados. A Figura 6.1 ilustra o comportamento da função de resposta ao item do MRG de quatro itens. A Figura 6.1(a) apresenta um item cuja categoria com maior escore(valor) apresenta um razoável valor para o parâmetro de dificuldade ( $b = 2.0$ ). A Figura 6.1(b) apresenta o mesmo item com o parâmetro de discriminação um pouco maior ( $a = 1,2$ ), o que faz com as curvas de todas as categorias tornem-se mais íngrimes. A Figura 6.1(c) apresenta o mesmo item com a categoria maior apresentando um parâmetro de dificuldade menor ( $b = 0,8$ ), o que faz com que a curva desta categoria se desloque para a esquerda, implicando que os indivíduos tenham uma chance razoável de atingi-la, com menor nível de habilidade. A Figura 6.1(d) apresenta o mesmo item com maior valor para o parâmetro de discriminação e menor para o parâmetro de dificuldade (da categoria maior) fazendo com que as curvas tornem-se mais íngrimes e, ao mesmo tempo, deslocando a curva da alternativa maior para a esquerda.

### 6.3 Estimação dos parâmetros dos itens

Descreveremos nas subseções seguintes o processo de estimação no **Modelo de Resposta Gradual**. Novamente faremos uso de notações e resultados obtidos dos Capítulos anteriores.

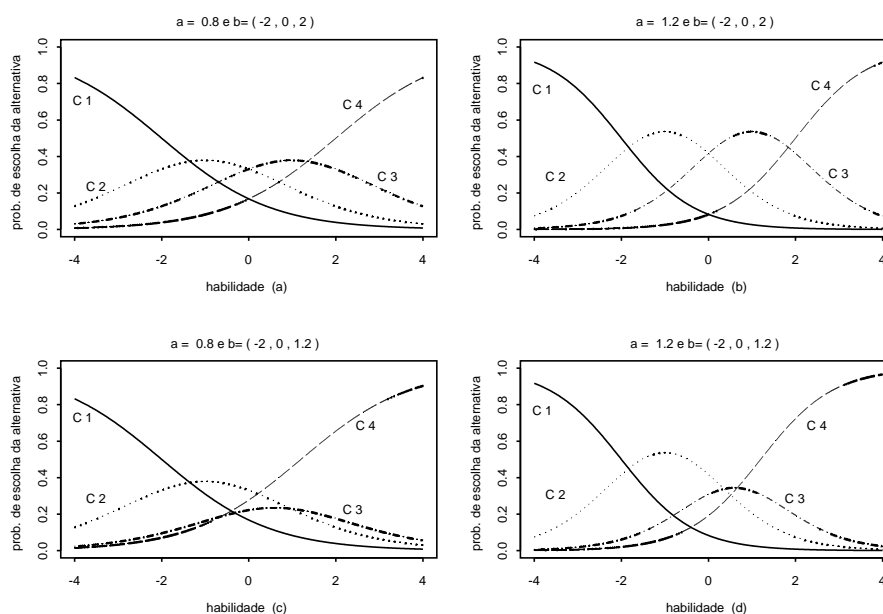


Figura 6.1 Representação gráfica do MRG de quatro itens

### 6.3.1 Habilidades conhecidas

Diferentemente do **Modelo de Resposta Nominal**, não precisaremos fazer restrições nos parâmetros para evitar problemas extras de indeterminação [Baker(1992)]. Portanto, com as suposições (S1), (S2) e (S3), temos que a verossimilhança é dada por

$$L(\zeta) = \prod_{i=1}^I \prod_{j=1}^n \prod_{s=0}^{m_i} P_{ijs}^{y_{ijs}},$$

com  $P_{ijs}$  como definido em (6.1), e a log-verossimilhança é dada por

$$l(\zeta) = \sum_{i=1}^I \sum_{j=1}^n \sum_{s=0}^{m_i} y_{ijs} \ln P_{ijs}. \quad (6.2)$$

O vetor escore, o qual fornece os emv dos parâmetros de interesse, é dado por

$$\begin{aligned} \mathbf{S}(\zeta_i) = \sum_{j=1}^n \frac{\partial}{\partial \zeta_i} \left\{ y_{ij0} \ln(1 - P_{ij1}^+) + y_{ij1} \ln(P_{ij1}^+ - P_{ij2}^+) + \dots + \right. \\ \left. y_{ij(h-1)} \ln(P_{ij(h-1)}^+ - P_{ijh}^+) + y_{ijh} \ln(P_{ijh}^+ - P_{ij(h+1)}^+) + \dots + \right. \\ \left. y_{ijm_i} \ln(P_{ijm_i}^+) \right\}. \end{aligned} \quad (6.3)$$

Detalhes do desenvolvimento pertinente a esta parte podem ser encontrados no Apêndice A.3. Por outro lado, temos que o vetor escore, para um determinado item e categoria, é dado por

$$\mathbf{S}(\zeta_{is}) = (S(a_i), S(b_{is}))' \quad (6.4)$$

Dessa forma, desenvolvendo (6.3) resulta em

$$S(a_i) = D \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right] \right\} \quad (6.5)$$

e

$$S(b_{is}) = Da_i \sum_{j=1}^n \left\{ W_{ijs}^+ \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} - \frac{y_{ijs}}{P_{ijs}} \right] \right\}. \quad (6.6)$$

que são as componentes de (6.4). Pela não-linearidade do sistema de equações descrito em (6.5) e (6.6) temos que usar algum método iterativo. Sendo assim, partindo de (6.4), temos que as componentes da matriz Hessiana são dadas por

$$\begin{aligned} H(a_i, a_i) = D^2 \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}^2} \left[ \left[ \left\{ (\theta_j - b_{is})^2 \{1 - 2P_{ijs}^+\} W_{ijs}^+ \right. \right. \right. \right. \\ \left. \left. \left. - (\theta_j - b_{i(s+1)})^2 \{1 - 2P_{ij(s+1)}^+\} W_{ij(s+1)}^+ \right\} \right] P_{ijs} \right. \\ \left. \left. - \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right]^2 \right] \right\}, \quad (6.7) \end{aligned}$$

$$\begin{aligned} H(b_{is}, b_{is}) = D^2 a_i^2 \sum_{j=1}^n W_{ijs}^+ \left\{ \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} - \frac{y_{ijs}}{P_{ijs}} \right] (1 - 2P_{ijs}^+) \right. \\ \left. - \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} + \frac{y_{ijs}}{P_{ijs}^2} \right] W_{ijs}^+ \right\}, \quad (6.8) \end{aligned}$$

$$H(b_{is}, b_{i(s-1)}) = D^2 a_i^2 \sum_{j=1}^n \left\{ \frac{y_{ij(s-1)} W_{ijs}^+ W_{ij(s-1)}^+}{P_{ij(s-1)}^2} \right\}, \quad (6.9)$$

$$H(b_{is}, b_{i(s+1)}) = D^2 a_i^2 \sum_{j=1}^n \left\{ \frac{y_{ij(s+1)} W_{ijs}^+ W_{ij(s+1)}^+}{P_{ij(s+1)}^2} \right\}, \quad (6.10)$$

$$\begin{aligned}
H(a_i, b_{is}) = & \sum_{j=1}^n \left\{ D \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} - \frac{y_{ijs}}{P_{ijs}} \right] W_{ijs}^+ \left\{ 1 + Da_i (\theta_j - b_{is}) (1 - 2P_{ijs}^+) \right\} \right. \\
& - D^2 a_i W_{ijs}^+ \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} \left[ (\theta_j - b_{i(s-1)}) W_{ij(s-1)}^+ - (\theta_j - b_{is}) W_{ijs}^+ \right] \right. \\
& \left. \left. - \frac{y_{ijs}}{P_{ijs}^2} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right] \right] \right\}, \quad (6.11)
\end{aligned}$$

com  $W_{ijs}^+ = P_{ijs}^+ Q_{ijs}^+$  e  $H(b_{is}, b_{i(s+d)}) = 0$  e  $H(b_{is}, b_{i(s-d)}) = 0$ ,  $d \geq 2$ . Dessa forma, temos que a matriz Hessiana, para um determinado item, pode ser escrita como

$$\mathbf{H}(\zeta_i) = \begin{bmatrix} H(b_{i1}, b_{i1}) & H(b_{i1}, b_{i2}) & 0 & \dots & \dots & H(b_{i1}, a_i) \\ H(b_{i2}, b_{i1}) & H(b_{i2}, b_{i2}) & H(b_{i2}, b_{i3}) & 0 & \dots & H(b_{i2}, a_i) \\ 0 & H(b_{i3}, b_{i2}) & H(b_{i3}, b_{i3}) & H(b_{i3}, b_{i4}) & \dots & H(b_{i3}, a_i) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & H(b_{im_i}, b_{im_i}) & H(b_{im_i}, a_i) \\ H(a_i, b_{i1}) & H(a_i, b_{i2}) & H(a_i, b_{i2}) & \dots & H(a_i, b_{im_i}) & H(a_i, a_i) \end{bmatrix}. \quad (6.12)$$

Tomando o valor esperado do simétrico de (6.12), teremos a Informação de Fisher

$$I(a_i, a_i) = D^2 \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{\left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right]^2}{P_{ijs}} \right\}, \quad (6.13)$$

$$I(b_{is}, b_{is}) = D^2 a_i^2 \sum_{j=1}^n \left\{ \left[ \frac{1}{P_{ij(s-1)}} + \frac{1}{P_{ijs}} \right] (W_{ijs}^+)^2 \right\}, \quad (6.14)$$

$$I(b_{is}, b_{i(s-1)}) = D^2 a_i^2 \sum_{j=1}^n \left\{ \frac{W_{ijs}^+ W_{ij(s-1)}^+}{P_{ij(s-1)}} \right\}, \quad (6.15)$$

$$I(b_{is}, b_{i(s+1)}) = D^2 a_i^2 \sum_{j=1}^n \left\{ \frac{W_{ijs}^+ W_{ij(s+1)}^+}{P_{ijs}} \right\} \quad (6.16)$$

e

$$\begin{aligned}
I(a_i, b_{is}) = & D^2 a_i \sum_{j=1}^n W_{ijs}^+ \left\{ \frac{1}{P_{ij(s-1)}} \left[ (\theta_j - b_{i(s-1)}) W_{ij(s-1)}^+ - (\theta_j - b_{is}) W_{ijs}^+ \right] \right. \\
& \left. - \frac{1}{P_{ijs}} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right] \right\}. \quad (6.17)
\end{aligned}$$

Dessa forma, os algoritmos de Newton-Raphson/Escore de Fisher podem ser empregados tais como descritos em (2.7) e (2.8).

### Habilidades conhecidas e agrupadas

Tal como descrito na Seção 4.3, temos que a log-verossimilhança é dada por

$$l(\Gamma_i) = const + \sum_{l=1}^q \sum_{s=0}^{m_i} y_{ils} \ln P_{ils}. \quad (6.18)$$

Diferenciando-se (6.18) com relação aos parâmetros dos itens, seguindo o mesmo desenvolvimento da abordagem de não-agrupamento, obtemos as componentes do vetor escore

$$S(a_i) = D \sum_{l=1}^q \sum_{s=0}^{m_i} \left\{ \frac{r_{ils}}{P_{ils}} \left[ (\bar{\theta}_l - b_{is}) W_{ils}^+ - (\bar{\theta}_l - b_{i(s+1)}) W_{il(s+1)}^+ \right] \right\} \quad (6.19)$$

e

$$S(b_{is}) = Da_i \sum_{l=1}^q \left\{ W_{ils}^+ \left[ \frac{r_{il(s-1)}}{P_{il(s-1)}} - \frac{r_{ils}}{P_{ils}} \right] \right\}, \quad (6.20)$$

em que  $P_{ils} \equiv P_{ijs}$ , com  $\theta_j$  substituído por  $\bar{\theta}_l$  e  $W_{ils}^+ = P_{ils}^+ Q_{ils}^+$ .

As componentes da matriz Hessiana são dadas por

$$\begin{aligned} H(a_i, a_i) = D^2 \sum_{l=1}^q \sum_{s=0}^{m_i} \left\{ \frac{r_{ils}}{P_{ils}^2} \left[ \left\{ (\bar{\theta}_l - b_{is})^2 \{1 - 2P_{ils}^+\} W_{ils}^+ \right. \right. \right. \\ \left. \left. - (\bar{\theta}_l - b_{i(s+1)})^2 \{1 - 2P_{il(s+1)}^+\} W_{il(s+1)}^+ \right\} P_{ils} \right. \\ \left. \left. - \left[ (\bar{\theta}_l - b_{is}) W_{ils}^+ - (\bar{\theta}_l - b_{i(s+1)}) W_{il(s+1)}^+ \right]^2 \right] \right\}, \quad (6.21) \end{aligned}$$

$$\begin{aligned} H(b_{is}, b_{is}) = D^2 a_i^2 \sum_{l=1}^q W_{ils}^+ \left\{ \left[ \frac{r_{il(s-1)}}{P_{il(s-1)}} - \frac{r_{ils}}{P_{ils}} \right] (1 - 2P_{ils}^+) \right. \\ \left. - \left[ \frac{r_{il(s-1)}}{P_{il(s-1)}^2} + \frac{r_{ils}}{P_{ils}^2} \right] W_{ils}^+ \right\}, \quad (6.22) \end{aligned}$$

$$H(b_{is}, b_{i(s-1)}) = D^2 a_i^2 \sum_{l=1}^q \left\{ \frac{r_{il(s-1)} W_{ils}^+ W_{il(s-1)}^+}{P_{il(s-1)}^2} \right\}, \quad (6.23)$$

$$H(b_{is}, b_{i(s+1)}) = D^2 a_i^2 \sum_{l=1}^q \left\{ \frac{r_{il(s+1)} W_{ils}^+ W_{il(s+1)}^+}{P_{il(s+1)}^2} \right\} \quad (6.24)$$

e

$$\begin{aligned} H(a_i, b_{is}) = & \sum_{j=1}^n \left\{ D \left[ \frac{r_{il(s-1)}}{P_{il(s-1)}} - \frac{r_{ils}}{P_{ils}} \right] W_{ils}^+ \{1 + Da_i (\bar{\theta}_l - b_{is}) (1 - 2P_{ils}^+)\} \right. \\ & - D^2 a_i W_{ils}^+ \left[ \frac{r_{il(s-1)}}{P_{il(s-1)}^2} [(\bar{\theta}_l - b_{i(s-1)}) W_{il(s-1)}^+ - (\bar{\theta}_l - b_{is}) W_{ils}^+] \right. \\ & \left. \left. - \frac{r_{ils}}{P_{ils}^2} [(\bar{\theta}_l - b_{is}) W_{ils}^+ - (\bar{\theta}_l - b_{i(s+1)}) W_{il(s+1)}^+] \right] \right\}. \end{aligned} \quad (6.25)$$

e as componentes da informação de Fisher dadas por

$$I(a_i, a_i) = D^2 \sum_{l=1}^q \sum_{s=0}^{m_i} \left\{ \frac{f_{il}}{P_{ils}} [(\bar{\theta}_l - b_{is}) W_{ils}^+ - (\bar{\theta}_l - b_{i(s+1)}) W_{il(s+1)}^+]^2 \right\}, \quad (6.26)$$

$$I(b_{ih}, b_{ih}) = D^2 a_i^2 \sum_{l=1}^q \left\{ f_{il} \left[ \frac{1}{P_{il(h-1)}} + \frac{1}{P_{ilh}} \right] (W_{ilh}^+)^2 \right\}, \quad (6.27)$$

$$I(b_{ih}, b_{i(h-1)}) = D^2 a_i^2 \sum_{l=1}^q \left\{ \frac{f_{il} W_{ilh}^+ W_{il(h-1)}^+}{P_{il(h-1)}} \right\}, \quad (6.28)$$

$$I(b_{ih}, b_{i(h+1)}) = D^2 a_i^2 \sum_{l=1}^q \left\{ \frac{f_{il} W_{ilh}^+ W_{il(h+1)}^+}{P_{ilh}} \right\} \quad (6.29)$$

e

$$\begin{aligned} I(a_i, b_{ih}) = & \sum_{l=1}^q f_{il} W_{ilh}^+ \left\{ \frac{1}{P_{il(h-1)}} [(\bar{\theta}_l - b_{i(h-1)}) W_{il(h-1)}^+ - (\bar{\theta}_l - b_{ih}) W_{ilh}^+] \right. \\ & \left. - \frac{1}{P_{ilh}} [(\bar{\theta}_l - b_{ih}) W_{ilh}^+ - (\bar{\theta}_l - b_{i(h+1)}) W_{il(h+1)}^+] \right\}. \end{aligned} \quad (6.30)$$

Na próxima Seção apresentaremos a estimação dos parâmetros dos itens ante ao desconhecimento das habilidades.

### 6.3.2 Habilidades desconhecidas

Temos que a probabilidade marginal, à semelhança do modelo de resposta nominal, é dada por

$$P(\mathbf{Y}_{.j}|\zeta, \eta) = \int_{\mathbb{R}} P(\mathbf{Y}_{.j}|\zeta, \theta) g(\theta|\eta) d\theta,$$

e a verossimilhança e log-verossimilhança dadas por

$$L(\zeta, \eta) = \prod_{j=1}^n P(\mathbf{Y}_{.j}|\zeta, \eta)$$

e

$$l(\zeta, \eta) = \sum_{j=1}^n \ln P(\mathbf{Y}_{.j}|\zeta, \eta). \quad (6.31)$$

As componentes do vetor escore, ver Apêndice A.3, em forma de pontos de quadratura, são dadas por

$$S(a_i) = D \sum_{l=1}^q \sum_{s=0}^{m_i} \left\{ \frac{\bar{r}_{ils}}{P_{ils}} \left[ (\bar{\theta}_l - b_{is}) W_{ils}^+ - (\bar{\theta}_l - b_{i(s+1)}) W_{il(s+1)}^+ \right] \right\} \quad (6.32)$$

e

$$S(b_{ih}) = Da_i \sum_{l=1}^q \left\{ W_{ilh}^+ \left[ \frac{\bar{r}_{il(h-1)}}{P_{il(h-1)}} - \frac{\bar{r}_{ilh}}{P_{ilh}} \right] \right\}. \quad (6.33)$$

com  $\bar{r}_{ils} = \sum_{j=1}^n y_{ijs} g_j^*(\bar{\theta}_l)$  e  $g_j^*(\bar{\theta}_l) = \frac{P(\mathbf{Y}_{.j}|\zeta, \bar{\theta}_l) g(\bar{\theta}_l|\eta)}{\sum_{l=1}^q P(\mathbf{Y}_{.j}|\zeta, \bar{\theta}_l) g(\bar{\theta}_l|\eta)}$ .

Samejima (1997), sugere uma adaptação do **algoritmo EM**, para a resolução das equações de estimação descritas em (6.32) e (6.33).

#### Adaptação do algoritmo EM

Basicamente, Samejima propôs uma extensão do uso do algoritmo EM nos modelos dicotômicos, com pequenas adaptações como descrito a seguir

#### Passo E

Usar os pontos de quadratura  $\bar{\theta}_l$ , os pesos  $A_l^{(t)}$ ,  $l = 1, \dots, q$  e estimativas provisórias dos parâmetros dos itens,  $\hat{\zeta}_i^{(t)}$ ,  $i = 1, \dots, I$ , para gerar  $g_j^*(\bar{\theta}_l)^{(t)}$  e, posteriormente,  $\bar{r}_{ils}^{(t)}$  e  $\bar{f}_{il}^{(t)}$ ,  $i = 1, \dots, I$ ,  $l = 1, \dots, q$  e  $s = 0, 1, \dots, m_i$ .



**Passo M**

com  $\bar{\mathbf{r}}^{(t)}$  e  $\bar{\mathbf{f}}^{(t)}$  obtidos no Passo E, resolver as equações de estimação para  $\zeta_i$ ,  $i = 1, \dots, I$ , usando o algoritmo de Newton-Raphson ou Escore de Fisher.

Como as quantidades  $\mathbf{r}$  e  $\mathbf{f}$  são substituídas pelos seus respectivos valores esperados  $\bar{\mathbf{r}}$  e  $\bar{\mathbf{f}}$ , os algoritmos iterativos podem ser implementados usando-se as expressões (6.21) a (6.30), com as convenientes substituições  $(\bar{\mathbf{r}}, \bar{\mathbf{f}})'$ .

Este procedimento gera estimativas consistentes dos parâmetros dos itens. A **superfície de verossimilhança** é **convexa** em quase toda parte. Dessa forma o vetor de **pontos estacionários** será único e portanto o máximo global da função de verossimilhança [Bock & Zimowski (1997)], [Samejima(1997)]. O processo de estimação convergirá a menos de padrões de resposta aberrantes [Bock & Zimowski (1997)]. A informação de Fisher fornecerá (desde que esta seja positiva definida) os erros-padrão assintóticos das estimativas dos parâmetros dos itens.

**6.4 Estimação das habilidades**

Usando o mesmo argumento dos modelos anteriores, temos que a verossimilhança para a habilidade de um determinado indivíduo é dada por

$$L(\theta_j) = \prod_{i=1}^I \prod_{s=0}^{m_i} P_{ijs}^{y_{ijs}},$$

e a log-verossimilhança dada por

$$l(\theta_j) = \sum_{i=1}^I \sum_{s=0}^{m_i} y_{ijs} \ln P_{ijs}, \quad (6.34)$$

gerando, ver desenvolvimentos no Apêndice A.3.2, o seguinte vetor escore

$$\begin{aligned} S(\theta_j) &= \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial \theta_j} \right) \right\} \\ &= D \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ \frac{a_i y_{ijs} (W_{ijs}^+ - W_{ij(s+1)}^+)}{P_{ijs}} \right\}. \end{aligned} \quad (6.35)$$

Como não é possível se obter solução explícita a partir da equação (6.35), temos que usar algum método iterativo. Para tanto temos que a matriz Hessiana é dada por

$$H(\theta_j) = D^2 \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ a_i^2 y_{ijs} \left[ \frac{\left( (1 - 2P_{ijs}^+) W_{ijs}^+ - (1 - 2P_{ij(s+1)}^+) W_{ij(s+1)}^+ \right)}{P_{ijs}} \right. \right. \\ \left. \left. - \frac{\left( W_{ijs}^+ - W_{ij(s+1)}^+ \right)^2}{P_{ijs}^2} \right] \right\}, \quad (6.36)$$

e a Informação de Fisher dada por

$$I(\theta_j) = D^2 a_i^2 \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{\left( W_{ijs}^+ - W_{ij(s+1)}^+ \right)^2}{P_{ijs}} \right\}. \quad (6.37)$$

Portanto, os métodos iterativos podem ser empregados como descritos em (2.32) e (2.33).

---

## Capítulo 7

# Modelo de Resposta Gradual: Estimação Bayesiana

---

### 7.1 Introdução

Abordaremos neste Capítulo a estimação Bayesiana no Modelo de Resposta Gradual.

### 7.2 Estimação dos parâmetros dos itens

Como descrito em (6.12), somente os parâmetros de dificuldade de categorias adjacentes são **assintoticamente dependentes**. Somado ao fato de que os mesmos podem assumir qualquer valor real, iremos propor uma distribuição **Normal Multivariada** como priori, ou seja,  $\mathbf{b}_i \sim N_{m_i}(\boldsymbol{\mu}_{\mathbf{b}_i}, \boldsymbol{\Sigma}_{\mathbf{b}_i})$ , a fim de se obter uma abordagem mais geral. Para os parâmetros de discriminação iremos propor uma distribuição log-normal, i. e.,  $a_i \sim \log - normal(\mu_{a_i}, \sigma_{a_i}^2)$ , pelos motivos explicitados na Seção 3.2.1. Ainda, consideraremos que os dois tipos de parâmetros (discriminação e dificuldade), são estocasticamente independentes entre si, em cada item. O mesmo vale para quaisquer parâmetros de diferentes itens.

#### 7.2.1 Habilidades conhecidas

Temos que a distribuição a posteriori pode ser escrita como

$$\begin{aligned} f(\boldsymbol{\zeta}|\mathbf{y}_{\dots}, \boldsymbol{\tau}, \boldsymbol{\theta}) &\propto L(\boldsymbol{\zeta}) f(\boldsymbol{\zeta}|\boldsymbol{\tau}) f(\boldsymbol{\tau}) \\ &= \left\{ \prod_{i=1}^I L(\zeta_i) \right\} \left\{ \prod_{i=1}^I f(\zeta_i|\tau_i) \right\} \left\{ \prod_{i=1}^I f(\tau_i) \right\}, \end{aligned} \quad (7.1)$$

e a log-verossimilhança como

$$\ln f(\boldsymbol{\zeta}|\mathbf{y}_{\dots}, \boldsymbol{\tau}, \boldsymbol{\theta}) \propto \sum_{i=1}^I l(\zeta_i) + \sum_{i=1}^I \ln f(\zeta_i|\tau_i) + \sum_{i=1}^I \ln f(\tau_i), \quad (7.2)$$

onde  $\boldsymbol{\tau}$  representa o vetor de hiperparâmetros associados a  $\boldsymbol{\zeta}$ . O vetor de **funções de estimação bayesianas** para os parâmetros de um determinado item é dado por

$$\mathbf{S}(\zeta_i)_B = \frac{\partial l(\zeta_i)}{\partial \zeta_i} + \frac{\partial \ln f(\zeta_i | \boldsymbol{\tau}_i)}{\partial \zeta_i}. \quad (7.3)$$

Segue-se de (6.5) e (3.5) que a função de estimação bayesiana, para o parâmetro de discriminação é dada por

$$S(a_i)_B = \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right] \right\} - \frac{1}{a_i} \left[ 1 + \frac{\ln a_i - \mu_{a_i}}{\sigma_{a_i}^2} \right]$$

e para os parâmetros de dificuldade de um determinado item, de (6.6) e (5.6) temos a seguinte função de estimação bayesiana

$$\mathbf{S}(\mathbf{b}_i)_B = \mathbf{S}(\mathbf{b}_i) - \mathbf{V}_{\mathbf{b}_i} (\mathbf{b}_i - \boldsymbol{\mu}_{\mathbf{b}_i}),$$

com  $\mathbf{S}(\mathbf{b}_i) = (S(b_{i1}), \dots, S(b_{im_i}))'$ . Novamente, estamos em face de um sistema não-linear, e, portanto, temos que aplicar algum método iterativo. Para tal, temos que, de (6.7) a (6.11), (3.10) e (5.6), a matriz Hessiana é dada por

$$\mathbf{H}(\zeta_i)_B = \mathbf{H}(\zeta_i) + \boldsymbol{\Lambda}_{\zeta_i}$$

e, de (6.13) a (6.17), (3.10) e (5.6), temos a informação de Fisher

$$\mathbf{I}(\zeta_i)_B = \mathbf{I}(\zeta_i) - \boldsymbol{\Lambda}_{\zeta_i},$$

com

$$\boldsymbol{\Lambda}_{\zeta_i} = \begin{bmatrix} \frac{[\sigma_{a_i}^2 + \ln a_i - \mu_{a_i} - 1]}{\mu_{a_i}^2 \sigma_{a_i}^2} & \mathbf{0} \\ \mathbf{0} & -\mathbf{V}_{\mathbf{b}_i} \end{bmatrix}. \quad (7.4)$$

Dessa forma, os procedimentos iterativos podem ser implementados tais como descritos em (2.7) e (2.8).

### Habilidades conhecidas e agrupadas

Usando (6.19) e (6.20), (6.21) a (6.25), (6.27) a (6.30), (5.5) e (5.6) temos que as expressões para o processo de estimação são dadas por

$$\mathbf{S}(\zeta_i)_B = \mathbf{S}(\zeta_i) + \boldsymbol{\lambda}_{\zeta_i},$$

$$\mathbf{H}(\zeta_i)_B = \mathbf{H}(\zeta_i) + \boldsymbol{\Lambda}_{\zeta_i}$$

e

$$\mathbf{I}(\zeta_i)_B = \mathbf{I}(\zeta_i) - \boldsymbol{\Lambda}_{\zeta_i} \quad (7.5)$$

onde  $\boldsymbol{\lambda}_{\zeta_i} = \left[ -\frac{1}{a_i} \left[ 1 + \frac{\ln a_i - \mu_{a_i}}{\sigma_{a_i}^2} \right]; -\mathbf{V}_{\mathbf{b}_i}(\mathbf{b}_i - \boldsymbol{\mu}_{\mathbf{b}_i}) \right]'$  e  $\boldsymbol{\Lambda}_{\zeta_i}$  como em (7.4).

Na próxima seção apresentaremos a estimação dos parâmetros dos itens mediante o desconhecimento das habilidades.

### 7.2.2 Habilidades desconhecidas

Seguindo o mesmo raciocínio usado para a estimação no Modelo de Resposta Nominal, temos que a **distribuição a posteriori marginal** é dada por

$$\begin{aligned} f(\boldsymbol{\psi}|\mathbf{y}_{\dots}) &\propto g(\boldsymbol{\eta}) \left\{ \int L(\mathbf{y}_{\dots}; \boldsymbol{\zeta}, \boldsymbol{\theta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \right\} \left\{ \int f(\boldsymbol{\zeta}|\boldsymbol{\tau}) f(\boldsymbol{\tau}) d\boldsymbol{\tau} \right\} \\ &= L(\mathbf{y}_{\dots}; \boldsymbol{\zeta}, \boldsymbol{\eta}) f(\boldsymbol{\zeta}) g(\boldsymbol{\eta}), \end{aligned} \quad (7.6)$$

onde  $\boldsymbol{\psi} = (\boldsymbol{\zeta}', \boldsymbol{\eta}')'$ . Por outro lado, o logaritmo de (7.6) é dado por

$$\begin{aligned} \ln f(\boldsymbol{\psi}|\mathbf{y}_{\dots}) &\propto \ln L(\boldsymbol{\zeta}, \boldsymbol{\eta}) + \ln f(\boldsymbol{\zeta}) + \ln g(\boldsymbol{\eta}) \\ &= l(\boldsymbol{\zeta}, \boldsymbol{\eta}) + \ln f(\boldsymbol{\zeta}) + \ln g(\boldsymbol{\eta}), \end{aligned} \quad (7.7)$$

e o vetor de **funções de estimação bayesianas** por

$$\begin{aligned} \mathbf{S}(\zeta_i)_{BM} &= \frac{\partial l(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} + \frac{\partial \ln f(\boldsymbol{\zeta})}{\partial \zeta_i} \\ &= \mathbf{S}(\zeta_i) + \boldsymbol{\lambda}_{\zeta_i}, \end{aligned} \quad (7.8)$$

em que  $\mathbf{S}(\zeta_i) = (S(a_i), \mathbf{S}(\mathbf{b}_i)')'$ , com  $S(a_i)$  e  $\mathbf{S}(\mathbf{b}_i) = (S(b_1), \dots, S(b_{m_i}))'$ , como dados em (6.5) e (6.6), respectivamente, e

$$\boldsymbol{\lambda}_{\zeta_i} = \left[ -\frac{1}{a_i} \left[ 1 + \frac{\ln a_i - \mu_{a_i}}{\sigma_{a_i}^2} \right]; [-\mathbf{V}_{\mathbf{b}_i}(\mathbf{b}_i - \boldsymbol{\mu}_{\mathbf{b}_i})] \right]'. \quad (7.9)$$

Mais uma vez, a não-linearidade do sistema de equações descrito em (7.8) nos leva a usar métodos iterativos. Dessa forma, de (6.21) a (6.30) e (7.4), temos que a matriz Hessiana e a informação de Fisher são dadas, respectivamente por

$$\mathbf{H}(\zeta_i)_{BM} = \mathbf{H}(\zeta_i) + \mathbf{\Lambda}_{\zeta_i}$$

e

$$\mathbf{I}(\zeta_i)_{BM} = \mathbf{I}(\zeta_i) - \mathbf{\Lambda}_{\zeta_i}.$$

Assim, o **algoritmo EM**, pode ser aplicado tal como descrito na Seção 6.3, usando-se a esperança da log-posteriori, equação (7.7), ao invés da log-verossimilhança, equação (6.31). Na próxima Seção apresentaremos a estimação das habilidades.

### 7.3 Estimação das habilidades

A distribuição a posteriori para uma determinada habilidade é dada por

$$g_j^*(\theta_j) = CP(\mathbf{Y}_{.j}|\theta_j, \zeta)g(\theta_j|\boldsymbol{\eta}) \quad (7.10)$$

$$\propto P(\mathbf{Y}_{.j}|\theta_j, \zeta)g(\theta_j|\boldsymbol{\eta}). \quad (7.11)$$

Novamente, adotaremos a média e a moda como estimadores dos traços latentes.

#### Estimação pela moda a posteriori

Tomando o logaritmo de (7.11), tem-se que

$$\ln g_j^*(\theta_j) = l_j^*(\theta_j) \propto l(\theta_j) + \ln g_j(\theta_j|\boldsymbol{\eta}). \quad (7.12)$$

Com a **função de estimação bayesiana**, dada por

$$S(\theta_j)_B = \frac{\partial l_j^*(\theta_j)}{\partial \theta_j} + \frac{\partial \ln g_j(\theta_j|\boldsymbol{\eta})}{\partial \theta_j}. \quad (7.13)$$

Notemos que o primeiro termo de (7.13) é tal como em (6.35), e adotando como priori a distribuição  $N(\mu_\theta, \sigma_\theta^2)$ , temos que (7.13) é dada por

$$S(\theta_j)_B = D \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ \frac{a_i y_{ijs} (W_{ijs}^+ - W_{ij(s+1)}^+)}{P_{ijs}} \right\} - \frac{\theta_j - \mu_\theta}{\sigma_\theta^2}. \quad (7.14)$$

Como a equação definida em (7.14) não possui solução explícita, novamente usaremos métodos iterativos. Portanto, de (6.36), (6.37) e (3.32) temos que a matriz Hessiana e a informação de Fisher, são dadas respectivamente por

$$H(\theta_j)_B = D^2 \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ a_i^2 y_{ijs} \left[ \frac{(1 - 2P_{ijs}^+) W_{ijs}^+ - (1 - 2P_{ij(s+1)}^+) W_{ij(s+1)}^+}{P_{ijs}} - \frac{(W_{ijs}^+ - W_{ij(s+1)}^+)^2}{P_{ijs}^2} \right] \right\} - \frac{1}{\sigma_\theta^2} \quad (7.15)$$

e

$$I(\theta_j)_B = D^2 \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ \frac{a_i^2 (W_{ijs}^+ - W_{ij(s+1)}^+)^2}{P_{ijs}} \right\} + \frac{1}{\sigma_\theta^2}. \quad (7.16)$$

Dessa forma, os métodos iterativos podem ser implementados tais como descritos em (2.32) e (2.33).

### Estimação pela média a posteriori

As expressões para a estimação pela média são idênticas aquelas para o Modelo de Resposta Nominal. A única diferença, é que nas expressões (5.22) e (5.23), a FRI é como dada em (6.1).

## 7.4 Comentários adicionais

Em geral, os procedimentos iterativos para a estimação dos parâmetros dos modelos policotômicos são mais lentos que aqueles para os modelos dicotômicos, muito embora, os problemas de convergência sejam menos freqüentes. Também necessitamos de um maior número de ciclos do **Algoritmo EM** para se obter a convergência dos processos iterativos [Van der Linden & Hambleton (1997)].

---

## Capítulo 8

# Outros Métodos de Estimação

---

### 8.1 Introdução

O objetivo deste Capítulo é o de apresentar métodos de estimação que visam possibilitar, de alguma forma, o aperfeiçoamento dos métodos descritos anteriormente. Serão descritos os métodos baseados em simulação de **Monte Carlo via Cadeias de Markov (MCMC)**, para os parâmetros dos itens e habilidades, e dois outros processos de estimação para as habilidades.

### 8.2 Estimação por Métodos de Monte Carlo via Cadeias de Markov

Basicamente, **Cadeias de Markov** compreendem o estudo de conjuntos de variáveis aleatórias que guardam estruturas de dependência entre si [Gamerman (1997)].

Os métodos de **Monte Carlo via Cadeias de Markov (MCMC)** simulam distribuições (em geral em contextos bayesianos) via cadeias de Markov construídas de modo conveniente. Essas cadeias são simuladas (dinamicamente) de modo a se atingir suas **distribuições estacionárias**, através das assim chamadas **distribuições condicionais completas** [Gilks, Richardson & Spiegelhalter (1996)].

Existem vários modos de se construir tais cadeias, porém, todos eles, incluindo o **amostrador de Gibbs** [Geman & Geman (1984)], são casos especiais de uma estrutura geral desenvolvida por Metropolis et al (1953) e Hastings (1970).

Particularmente, no contexto da TRI, Patz & Junker (1999a) propuseram um mecanismo de simulação baseado no então conhecido algoritmo de **Metropolis - Hastings dentro do Amostrador de Gibbs (MHDAG)** [Tierney (1994)]. Este processo simula valores iterativamente das distribuições condicionais completas de acordo com o **algoritmo de Gibbs** [Gamerman (1997)], porém com as referidas distribuições que possuem forma conhecida a menos de uma constante de normalização, para qual usa-se uma iteração simples do algoritmo de **Metropolis-Hastings**. [Gamerman (1997)]

Sucintamente, o algoritmo MHDAG pode ser descrito do seguinte modo. Considere  $q_{\theta}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)})$  e  $q_{\zeta}(\boldsymbol{\zeta}^{(t)}, \boldsymbol{\zeta}^{(t+1)})$  (com  $t$  denotando iteração) as densidades propostas relativas as habilidades e aos parâmetros dos itens, respectivamente. Para cada passo da transição do algoritmo temos

1. Simular  $\boldsymbol{\theta}^{(t)} \sim g(\boldsymbol{\theta}|\boldsymbol{\zeta}^{(t-1)}, \mathbf{y})$  (condicional completa) através de :



- (a) Simular  $\theta^{(*)} \sim q_{\theta}(\theta^{(t-1)}, \theta)$   
 (b) Aceitar  $\theta^{(t)} = \theta^{(*)}$  com probabilidade

$$\alpha(\theta^{(t)}, \theta^{(*)}) = \min \left\{ \frac{L(\zeta^{(t-1)}, \theta^{(*)})g(\theta^{*}, \zeta^{(t-1)})q_{\theta}(\theta^{*}, \theta^{(t-1)})}{L(\zeta^{(t-1)}, \theta^{(t-1)})g(\theta^{(t-1)}, \zeta^{(t-1)})q_{\theta}(\theta^{(t-1)}, \theta^{*})}, 1 \right\}$$

Caso contrário, faça  $\theta^{(t)} = \theta^{(t-1)}$

2. Simular  $\zeta^{(t)} \sim g(\zeta|\theta^{(t)}, \mathbf{y})$  (condicional completa) através de

- (a) Simular  $\zeta^{(*)} \sim q_{\zeta}(\zeta^{(t-1)}, \zeta)$   
 (b) Aceitar  $\zeta^{(t)} = \zeta^{(*)}$  com probabilidade

$$\alpha(\zeta^{(t)}, \zeta^{(*)}) = \min \left\{ \frac{L(\zeta^{*}, \theta^{(t)})g(\theta^{(t)}, \zeta^{(*)})q_{\zeta}(\zeta^{*}, \zeta^{(t-1)})}{L(\zeta^{(t-1)}, \theta^{(t)})g(\theta^{(t)}, \zeta^{(t-1)})q_{\zeta}(\zeta^{(t-1)}, \zeta^{*})}, 1 \right\}$$

Caso contrário, faça  $\zeta^{(t)} = \zeta^{(t-1)}$

com  $g(\theta, \zeta)$  denotando uma priori conveniente (neste caso conjunta) e  $L(\theta, \zeta)$  uma verossimilhança genuína, e.g., equação(2.2)(modelos dicotômicos).

Uma das grandes vantagens desse método é que, além de ser capaz de acoplar de modo relativamente simples modelos mais complexos, e.g., **modelos policotômicos e observações faltantes** [Patz & Junker (1999b), van der Linden & Hambleton (1997), Boomsma, Van Duijn & Snijders (2001)], eles fornecem uma aproximação empírica das verdadeiras distribuições (a posteriori) dos parâmetros de interesse [Gilks, Richardson & Spiegelhalter (1996)], ao invés de meramente estimativas pontuais e erros-padrão, como na abordagem de Bock & Aitkin e a de Mislevy.

Nas próximas duas subseções exemplificaremos os algoritmos para os modelos abordados neste trabalho. As notações, suposições e prioris são as mesmas dos capítulos anteriores.

### 8.2.1 Modelos Dicotômicos

O algoritmo pode ser descrito como

1. Simular  $\theta_j^{(t)} \sim g(\theta_j|\zeta^{(t-1)}, \mathbf{y}_j)$ (condicional completa), para  $j = 1, \dots, n$  independentemente, considerando como priori uma  $N(\theta_j|\mu_{\theta_j}, \sigma_{\theta_j}^2)$  através de :

- (a) Simular  $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \psi_{\theta_j})$   
 (b) Calcular o vetor de probabilidades de aceitação  $\theta_j^{(t)} = \theta_j^{(*)}$

$$\pi_j(\theta_j^{(t-1)}, \theta_j^{(*)}) = \min \left\{ \frac{L(\zeta^{(t-1)}, \theta_j^{(*)}) \exp \left\{ -\frac{(\theta_j^{(*)} - \mu_{\theta_j})^2}{2\sigma_{\theta_j}^2} \right\}}{L(\zeta^{(t-1)}, \theta_j^{(t-1)}) \exp \left\{ -\frac{(\theta_j^{(t-1)} - \mu_{\theta_j})^2}{2\sigma_{\theta_j}^2} \right\}}, 1 \right\}$$

- (c) Aceitar cada  $\theta_j^{(t)} = \theta_j^{(*)}$  com probabilidade  $\pi_j$ , caso contrário  $\theta_j^{(t)} = \theta_j^{(t-1)}$
2. Simular  $\zeta_i^{(t)} \sim g(\zeta_i | \boldsymbol{\theta}^{(t)}, \mathbf{y}_{..})$  (condicional completa), para  $i = 1, \dots, I$  independentemente, considerando como priori  $f(\zeta_i | \boldsymbol{\tau}_i) \equiv \log\text{-normal}(a_i | \mu_{a_i}, \sigma_{a_i}^2) \times N(b_i | \mu_{b_i}, \sigma_{b_i}^2) \times \text{Beta}(c_i | \alpha_i - 1, \beta_i - 1)$  através de :

- (a) Simular  $a_i^{(*)} \sim \log\text{-normal}(a_i^{(*)} | a_i^{(t-1)}, \psi_{a_i})$ ,  $b_i^{(*)} \sim N(b_i^{(*)} | b_i^{(t-1)}, \psi_{b_i})$  e  $c_i^{(*)} \sim \text{Beta}(c_i^{(*)} | s_i^{(t-1)} - 1, r_i^{(t-1)} - 1)$

- (b) Calcular o vetor de probabilidades de aceitação  $\zeta^{(t)} = \zeta^{(*)}$

$$\pi_i(\zeta^{(t-1)}, \zeta^{(*)}) = \min\{R_{\zeta_i}, 1\}$$

com

$$\begin{aligned} R_{\zeta_i} &= \frac{L(\zeta_i^{(*)}, \boldsymbol{\theta}^{(t)}) \exp\left\{-\frac{(\ln a_i^{(*)} - \mu_{a_i})^2}{2\sigma_{a_i}^2}\right\} \exp\left\{-\frac{(b_i^{(*)} - \mu_{b_i})^2}{2\sigma_{b_i}^2}\right\}}{L(\zeta_i^{(t-1)}, \boldsymbol{\theta}^{(t)}) \exp\left\{-\frac{(\ln a_i^{(t-1)} - \mu_{a_i})^2}{2\sigma_{a_i}^2}\right\} \exp\left\{-\frac{(b_i^{(t-1)} - \mu_{b_i})^2}{2\sigma_{b_i}^2}\right\}} \\ &\times \frac{[c_i^{(*)}]^{\alpha_i - 2 + s_i^{(t-1)} - 1} [a_i^{(*)}]^2 \exp\left\{-\frac{(\ln a_i^{(*)} - a_i^{(t-1)})^2}{2\psi_{a_1}^2}\right\}}{[c_i^{(t-1)}]^{\alpha_i - 2 + s_i^{(*)} - 1} [a_i^{(t-1)}]^2 \exp\left\{-\frac{(\ln a_i^{(t-1)} - a_i^{(*)})^2}{2\psi_{a_1}^2}\right\}} \\ &\times \frac{\beta(s_i^{(*)} - 1, r_i^{(*)} - 1) [1 - c_i^{(*)}]^{\beta_i - 2 + r_i^{(t-1)} - 1}}{\beta(s_i^{(t-1)} - 1, r_i^{(t-1)} - 1) [1 - c_i^{(t-1)}]^{\beta_i - 2 + r_i^{(*)} - 1}} \end{aligned}$$

- (c) Aceitar cada  $\zeta_i^{(t)} = \zeta_i^{(*)}$  com probabilidade  $\pi_i$ , caso contrário  $\zeta_i^{(t)} = \zeta_i^{(t-1)}$

### 8.2.2 Modelos Policotômicos

#### Modelo de Resposta Nominal

1. Simular  $\theta_j^{(t)} \sim g(\theta_j | \boldsymbol{\Gamma}^{(t-1)}, \mathbf{y}_{...})$  (condicional completa), para  $j = 1, \dots, n$  independentemente, considerando como priori uma  $N(\theta_j | \mu_{\theta_j}, \sigma_{\theta_j}^2)$  através de :
- (a) Simular  $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \psi_{\theta_j})$
- (b) Calcular o vetor de probabilidades de aceitação  $\theta_j^{(t)} = \theta_j^{(*)}$

$$\pi_j \left( \theta_j^{(t-1)}, \theta_j^{(*)} \right) = \min \left\{ \frac{L(\boldsymbol{\Gamma}^{(t-1)}, \theta_j^{(*)}) \exp \left\{ -\frac{(\theta_j^{(*)} - \mu_{\theta_j})^2}{2\sigma_{\theta_j}^2} \right\}}{L(\boldsymbol{\Gamma}^{(t-1)}, \theta_j^{(t-1)}) \exp \left\{ -\frac{(\theta_j^{(t-1)} - \mu_{\theta_j})^2}{2\sigma_{\theta_j}^2} \right\}}, 1 \right\}$$

(c) Aceitar cada  $\theta_j^{(t)} = \theta_j^{(*)}$  com probabilidade  $\pi_j$ , caso contrário  $\theta_j^{(t)} = \theta_j^{(t-1)}$

2. Simular  $\boldsymbol{\Gamma}_i^{(t)} \sim g(\boldsymbol{\Gamma}_i | \boldsymbol{\theta}^{(t)}, \mathbf{y}_{\dots})$  (condicional completa), para  $i = 1, \dots, I$  independentemente, considerando como priori  $f(\boldsymbol{\Gamma}_i | \boldsymbol{\tau}_i) \equiv N_{2m_i-2}(\boldsymbol{\mu}_{\boldsymbol{\Gamma}_i}, \boldsymbol{\Sigma}_i)$  através de

(a) Simular  $\boldsymbol{\Gamma}_i^{(*)} \sim N_{2m_i-2}(\boldsymbol{\Gamma}_i^{(t-1)}, \boldsymbol{\Psi}_i)$

(b) Calcular o vetor de probabilidades de aceitação  $\boldsymbol{\Gamma}_i^{(t)} = \boldsymbol{\Gamma}_i^{(*)}$

$$\pi_i \left( \boldsymbol{\Gamma}_i^{(t-1)}, \boldsymbol{\Gamma}_i^{(*)} \right) = \min \{ R_{\boldsymbol{\Gamma}_i}, 1 \}$$

com

$$R_{\boldsymbol{\Gamma}_i} = \frac{L(\boldsymbol{\Gamma}_i^{(*)}, \boldsymbol{\theta}^{(t)}) \exp \left\{ -\frac{1}{2} \left[ \left( \boldsymbol{\Gamma}_i^{(*)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i} \right)' \boldsymbol{\Sigma}_i^{-1} \left( \boldsymbol{\Gamma}_i^{(*)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i} \right) \right] \right\}}{L(\boldsymbol{\Gamma}_i^{(t-1)}, \boldsymbol{\theta}^{(t)}) \exp \left\{ -\frac{1}{2} \left[ \left( \boldsymbol{\Gamma}_i^{(t-1)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i} \right)' \boldsymbol{\Sigma}_i^{-1} \left( \boldsymbol{\Gamma}_i^{(t-1)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i} \right) \right] \right\}}$$

(c) Aceitar cada  $\boldsymbol{\Gamma}_i^{(t)} = \boldsymbol{\Gamma}_i^{(*)}$  com probabilidade  $\pi_i$ , caso contrário  $\boldsymbol{\Gamma}_i^{(t)} = \boldsymbol{\Gamma}_i^{(t-1)}$

### Modelo de Resposta Gradual

1. Simular  $\theta_j^{(t)} \sim g(\theta_j | \boldsymbol{\zeta}^{(t-1)}, \mathbf{y})$  (condicional completa), para  $j = 1, \dots, n$  independentemente, considerando como priori uma  $N(\theta_j | \mu_{\theta_j}, \sigma_{\theta_j}^2)$  através de :

(a) Simular  $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \psi_{\theta_j})$

(b) Calcular o vetor de probabilidades de aceitação  $\theta_j^{(t)} = \theta_j^{(*)}$

$$\pi_j \left( \theta_j^{(t-1)}, \theta_j^{(*)} \right) = \min \left\{ \frac{L(\boldsymbol{\zeta}^{(t-1)}, \theta_j^{(*)}) \exp \left\{ -\frac{(\theta_j^{(*)} - \mu_{\theta_j})^2}{2\sigma_{\theta_j}^2} \right\}}{L(\boldsymbol{\zeta}^{(t-1)}, \theta_j^{(t-1)}) \exp \left\{ -\frac{(\theta_j^{(t-1)} - \mu_{\theta_j})^2}{2\sigma_{\theta_j}^2} \right\}}, 1 \right\}$$

(c) Aceitar cada  $\theta_j^{(t)} = \theta_j^{(*)}$  com probabilidade  $\pi_j$ , caso contrário  $\theta_j^{(t)} = \theta_j^{(t-1)}$

2. Simular  $\zeta_i^{(t)} \sim g(\zeta_i | \theta_i^{(t)}, \mathbf{y})$  (condicional completa), para  $i = 1, \dots, I$  independentemente, considerando como priori  $f(\zeta_i | \tau_i) \equiv \log - normal(a_i | \mu_{a_i}, \sigma_{a_i}^2) \times N_{m_i}(\mathbf{b}_i | \mu_{\mathbf{b}_i}, \Sigma_{\mathbf{b}_i})$  através de:

- (a) Simular  $a_i^{(*)} \sim \log - normal(a_i^{(*)} | a_i^{(t-1)}, \psi_{a_i})$ , e  $\mathbf{b}_i^{(*)} \sim N_{m_i}(\mathbf{b}_i^{(*)} | \mathbf{b}_i^{(t-1)}, \Psi_{\mathbf{b}_i})$   
 (b) Calcular o vetor de probabilidades de aceitação  $\zeta_i^{(t)} = \zeta_i^{(*)}$

$$\pi_i(\zeta^{(t-1)}, \zeta^{(*)}) = \min \{R_{\zeta_i}, 1\}$$

com

$$R_{\zeta_i} = \frac{L(\zeta_i^{(*)}, \theta^{(t)}) \exp \left\{ -\frac{(\ln a_i^{(*)} - \mu_{a_i})^2}{2\sigma_{a_i}^2} \right\} \exp \left\{ -\frac{(\ln a_i^{(*)} - a_i^{(t-1)})^2}{2\psi_{a_i}} \right\} [a_i^{(*)}]^2}{L(\zeta_i^{(t-1)}, \theta^{(t)}) \exp \left\{ -\frac{(\ln a_i^{(t-1)} - \mu_{a_i})^2}{2\sigma_{a_i}^2} \right\} \exp \left\{ -\frac{(\ln a_i^{(t-1)} - a_i^{(*)})^2}{2\psi_{a_i}} \right\} [a_i^{(t-1)}]^2} \times \frac{\exp \left\{ -\frac{1}{2} \left[ (\mathbf{b}_i^{(*)} - \mu_{\mathbf{b}_i})' \Sigma_{\mathbf{b}_i}^{-1} (\mathbf{b}_i^{(*)} - \mu_{\mathbf{b}_i}) \right] \right\}}{\exp \left\{ -\frac{1}{2} \left[ (\mathbf{b}_i^{(t-1)} - \mu_{\mathbf{b}_i})' \Sigma_{\mathbf{b}_i}^{-1} (\mathbf{b}_i^{(t-1)} - \mu_{\mathbf{b}_i}) \right] \right\}}$$

- (c) Aceitar cada  $\zeta_i^{(t)} = \zeta_i^*$  com probabilidade  $\pi_i$ , caso contrário  $\zeta_i^{(t)} = \zeta_i^{(t-1)}$

Na próxima Seção apresentaremos alguns processos de estimação das habilidades.

### 8.3 Estimação por Máxima Verossimilhança Ponderada

Apesar de os emv das habilidades serem consistentes, eles são viciados, e muito embora, os estimadores **MAP** e **EAP**, apresentam vícios relativamente menores (notadamente este último), estes vícios podem apresentar certas magnitude para testes pequenos (poucos itens).

Lord (1983) alerta que os vícios podem trazer sérios problemas, como por exemplo nos **processos de equalização** [Andrade (2001)]. Em vista disso, Warm (1989), propôs um método de estimação que diminuiu consideravelmente o vício associado [de  $O(n^{-1})$  para  $o(n^{-1})$ , Sen & Singer (1993)]. A idéia básica foi considerar uma **função ponderadora** (peso) e inseri-la na verossimilhança (como uma priori), e então maximizar essa nova função com relação as habilidades. Daí então a denominação de **estimador ponderado de máxima verossimilhança (EPMV)**.

Consideremos então a seguinte função (ponderada)

$$\begin{aligned} L_W(\boldsymbol{\theta}) &= L(\boldsymbol{\theta}) w(\boldsymbol{\theta}) \\ &= \prod_{j=1}^n \{L(\theta_j) w(\theta_j)\}, \end{aligned}$$

em que  $L(\theta_j)$  é uma verossimilhança genuína (das habilidades) e  $w(\theta_j)$  é uma função ponderadora.

Baseado nos resultados de Lord (1984), que estabeleceu uma relação entre os vícios dos estimadores MV e MAP das habilidades, Warm (1989) propôs uma função peso de sorte que

$$\frac{\partial \ln w(\theta_j)}{\partial \theta_j} = \frac{1}{2} I^{-1}(\theta_j) J(\theta_j), \quad (8.1)$$

onde  $I(\theta_j)$  é a informação de Fisher associada a habilidade do  $j$ -ésimo indivíduo e

$$J(\theta_j) = \sum_{i=1}^I \left\{ \frac{P_{ij}^{(1)} P_{ij}^{(2)}}{P_{ij} Q_{ij}} \right\}, \quad (8.2)$$

com

$$P_{ij}^{(d)} = \frac{\partial^{(d)} P_{ij}}{\partial \theta_j^{(d)}}.$$

Muito embora Kim & Nicewander (1993) tenham sugerido que o EPMV possa ser visto como um método bayesiano utilizando a priori de Jefferys, Warm (1989) adverte que isso só é válido para os MLUP1 e MLUP2 e, mesmo assim, somente de um ponto de vista matemático, haja vista que as idéias que estão embutidas em cada um dos métodos são bem distintas. Enquanto que no primeiro, a idéia é introduzir alguma informação (subjéitiva) no processo de estimação, o segundo tenta extrair da própria amostra (teste) informações que possibilitem diminuir o viés associado as estimativas.

Dessa forma de (8.1) e (8.2), temos que a equação de estimação ponderada é dada por

$$S_w(\theta_j) = S(\theta_j) + \frac{1}{2} I^{-1}(\theta_j) J(\theta_j). \quad (8.3)$$

Warm conseguiu demonstrar que, muito embora (8.3) seja uma função de estimação viciada [Jorgensen & Laboriau (1992)], ela produz estimadores com vieses reduzidos em relação aos métodos convencionais.

## 8.4 Estimação por Máxima Verossimilhança Bi-Ponderada

O processo de estimação aqui descrito se aplica, somente, aos modelos dicotômicos. As expressões aqui apresentadas servem para o modelo de 3 parâmetros, e podem ser simplificadas para os outros dois modelos dicotômicos referidos neste trabalho.

Com o objetivo de tentar contornar problemas de (como são conhecidos na linguagem educacional) flutuações na chance de acerto casual (que pode levar indivíduos com baixa habilidade a responder corretamente a um grande número de itens difíceis) e displicência por parte dos respondentes (que pode acarretar que indivíduos com habilidades alta respondam incorretamente um grande número de itens fáceis), Mislevy & Bock (1982) propuseram a **estimação bi-ponderada das habilidades**.

Os emv dos traços latentes são bastante sensíveis aos problemas descritos anteriormente, mesmo quando estes não são tão presentes nos testes [Mislevy & Bock (1982)].

Basicamente, Mislevy & Bock adaptaram a idéia que Mosteller & Tukey (1977) sugeriram para estimação **robusta** da média. A proposta é modificar a equação de verossimilhança da seguinte forma :

$$S(\theta_j) = \sum_{i=1}^I V_{ij} W_{ij} a_i (1 - c_i) (y_{ij} - P_{ij}) = 0,$$

com

$$V_{ij} = \begin{cases} (1 - U_{ij}^2)^2, & \text{se } |U_{ij}| \leq 1 \\ 0, & \text{caso contrário} \end{cases}$$

e  $U_{ij} = \frac{b_i - \theta_j}{\omega \alpha_i}$ , onde  $\alpha_i = a_i^{-1}$  é conhecido como **dispersão** do item e  $\omega$  é uma constante arbitrária. Quanto menor o valor dessa constante, menor será o peso atribuído. A idéia é portanto, dar diferentes pesos aos itens no processo de estimação, levando em consideração seus parâmetros.

---

## Capítulo 9

# Recursos Computacionais e Simulação

---

### 9.1 Introdução

Nesse Capítulo discutiremos alguns aspectos computacionais relativos aos modelos apresentados. Subseqüentemente, com o intuito de ilustrar alguns dos **procedimentos de estimação** apresentados nos capítulos anteriores, conduziremos dois estudos de simulação. O primeiro concerne ao MRN (Capítulos 4 e 5) e o segundo concentra-se numa comparação entre os MLUP (Capítulos 2 e 3) e o MRN. Os programas utilizados bem como os respectivos bancos de dados gerados podem ser solicitados via e-mail : *c\_naber@yahoo.com.br*.

### 9.2 Recursos Computacionais

Sem dúvida, a grande difusão do uso da TRI está intimamente ligada ao avanço dos programas computacionais desenvolvidos ao longo dos anos. Isto possibilitou o uso dessa modelagem em grandes sistemas de avaliações educacionais. Nas duas próximas subseções abordaremos de modo sucinto alguns aspectos computacionais.

#### 9.2.1 Modelos Dicotômicos

Para o ajuste de modelos dicotômicos há vários programas. Os dois mais utilizados no Brasil, são **BILOG-G** [Mislevy & Bock (1990)] e **BILOG-MG** [Zimowski et al (1996)]. Neles estão implementados, respectivamente, modelos para uma e para mais de uma população. Ambos se encontram disponíveis para a plataforma **Windows**. No que concerne a estimação dos parâmetros dos itens, há a opção de **Máxima Verossimilhança Marginal** via algoritmo EM [Bock & Aitkin (1981)] e **Moda Marginal a Posteriori** [Mislevy (1986)]. Com relação às habilidades, estão disponíveis os métodos de **Máxima Verossimilhança**, **Esperança a Posteriori** e **Moda a Posteriori** [Baker (1992)].

Também estão disponíveis os procedimentos iterativos de **Newton - Raphson** e **Escore de Fisher**. Além destas, existem outras opções como **Estimação das Densidades Latentes** [Mislevy (1984)] **Acelerador de Ramsey** [Ramsey (1979)] dentre outras.

### 9.2.2 Modelos Policotômicos

Os dois principais programas para a análise de modelos policotômicos são o **Multilog** [Thissen (1991)] e **Parscale** [Muraki & Bock (1997)]. No primeiro encontram-se implementados os modelos dicotômicos (1, 2 e 3 parâmetros), o **modelo de Resposta Nominal**, o **Modelo de Resposta Gradual** e o **Modelo de Múltipla Escolha** [Thissen & Steinberg (1984)]. No segundo estão implementados os **Modelos de Resposta Gradual** e os **Modelos de Créditos Parciais**. As versões mais recentes de ambos permitem a análise na presença de vários grupos. As opções de estimação são basicamente as mesmas dos programas de modelos dicotômicos. Ambos já se encontram para a plataforma **Windows**. Maiores informações sobre os programas descritos podem ser encontradas em *www.ssicentral.com*.

## 9.3 Simulações

Nesta parte apresentamos algumas simulações realizadas com o intuito de comparar métodos de estimação e modelos de resposta ao item.

### 9.3.1 Modelo de Resposta Nominal

Nesta seção conduziremos comparações entre os métodos de estimação Máxima Verossimilhança e Bayesiano no Modelo de Resposta Nominal. Considerou-se um conjunto de 5000 valores gerados a partir de uma distribuição  $N(0,1)$  (representando as habilidades dos indivíduos), e um conjunto de 50 itens de 5 alternativas, com parâmetros que variam de  $a$  (discriminação) : -1,10 a 1,10 e  $b$  (dificuldade) : -3,00 a 3,50. Esses valores encontram-se na Tabela B.1 do Apêndice B. Estes são considerados valores razoáveis para os parâmetros dos itens, no sentido de que tenhamos itens que variam em dificuldade e discriminação (para maior entendimento vide Seção 4.2). A partir desses valores geramos um conjunto de respostas policotômicas (vide Davey, Nering & Thompson (1997) e Azevedo (2002), para maiores detalhes). Usamos os métodos de MVM e BM (Capítulos 4 e 5) para obter estimativas dos parâmetros dos itens. De posse de cada um desses conjuntos de valores obtivemos estimativas para as habilidades usando cada um dos três métodos expostos neste trabalho (Capítulos 4 e 5). Para a geração das respostas, bem como para a estimação dos parâmetros usou-se programas construídos via linguagem OX versão 3.2 [Doornik (2000), este manual é relativo à versão 3.0]. Em todos os resultados dos parâmetros dos itens estamos considerando os parâmetros irrestritos (vide Seção 4.3).

Nesta parte apresentaremos os resultados pertinentes aos parâmetros dos itens. Deixaremos os resultados das habilidades para a próxima seção. Para a estimação Bayesiana adotou-se uma priori  $N(0,1)$  para as habilidades e  $N(0,16)$  para os parâmetros dos itens. Estes são valores sugeridos pela literatura [Thissen (1991)] e que produziram bons resultados. Uma justificativa para a variância da priori dos parâmetros dos itens é que na escala  $(0,1)$  é esperado que tenhamos



itens com parâmetros de dificuldade que variem de (-3,3). E como a distribuição adotada é a Normal, um valor igual a 4 para o desvio-padrão é suficiente para cobrir o intervalo dos valores dos parâmetros.

As estatísticas usadas na comparação dos métodos foram as seguintes:

- Soma dos Quadrados dos Resíduos (Vícios) (SQR) : é o quadrado da diferença entre o verdadeiro valor do parâmetro (item ou habilidade) e sua estimativa.

(item)

$$\left( \sum_{i=1}^I (\hat{\alpha}_i - \alpha_i)^2 \quad , \quad \sum_{i=1}^I (\hat{\delta}_i - \delta_i)^2 \right)$$

(habilidade)

$$\left( \sum_{j=1}^n (\hat{\theta}_j - \theta_j)^2 \right)$$

- Soma das Variâncias associadas às estimativas dos parâmetros (SQVar): É o elemento da diagonal da inversa da matriz Hessiana (Informação de Fisher).
- Soma dos Erros Quadráticos Médios associados às estimativas (SEQM) : é a soma da variância com o vício ao quadrado (quadrado do resíduo) associado às estimativas.

Da Tabela 9.1 notamos um desempenho equivalente dos dois métodos e também um excelente desempenho dos estimadores. As estimativas mostraram-se bem próximas dos verdadeiros valores e os erros-padrão apresentaram baixa magnitude.

Tabela 9.1 *Estatísticas relativas às estimativas dos parâmetros dos itens*

Estatística	Parâmetros Irrestritos			
	Discr. ( $\alpha$ )		Difc. ( $\delta$ )	
	MVM	BM	MVM	BM
<b>SQR</b>	1,201	1,211	1,438	1,439
<b>SQVar</b>	1,110	1,102	1,418	1,407
<b>SEQM</b>	2,311	2,313	2,856	2,847

A Figura 9.1 ilustra o excelente desempenho de ambos os tipos de estimadores (**máxima verossimilhança** e **bayesianos**). Notamos que as estimativas mostram-se relativamente próximas dos verdadeiros valores. A Figura 9.2 sugere uma equivalência nas estimativas pontuais (2 primeiros gráficos) e nos erros-padrão (2 últimos gráficos) das estimativas por ambos os métodos. Essa equivalência pode ser devida a, por exemplo, a quantidade de dados e a ausência do parâmetro de acerto ao acaso neste modelo. Este parâmetro, em geral, provoca instabilidade nos processos de estimação.

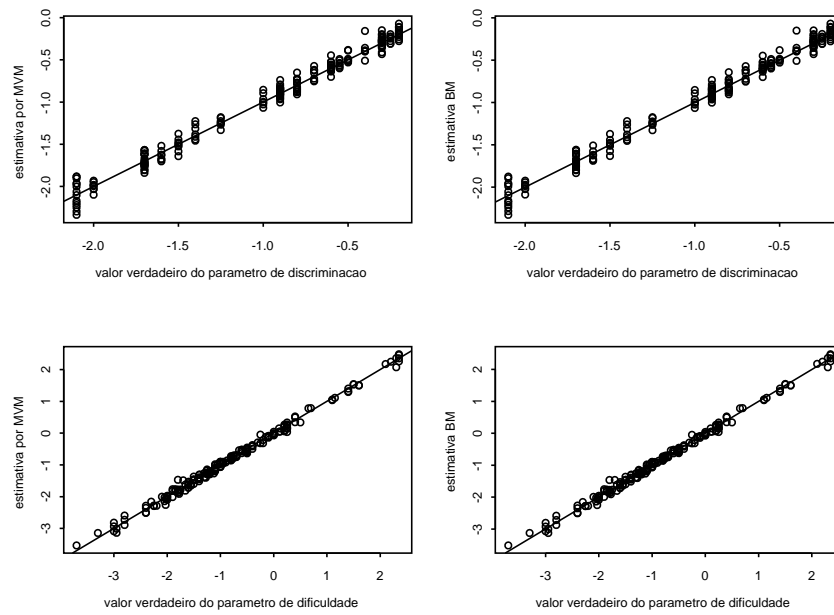


Figura 9.1 *Dispersão entre os verdadeiros parâmetros dos itens e as estimativas - parâmetros irrestritos*

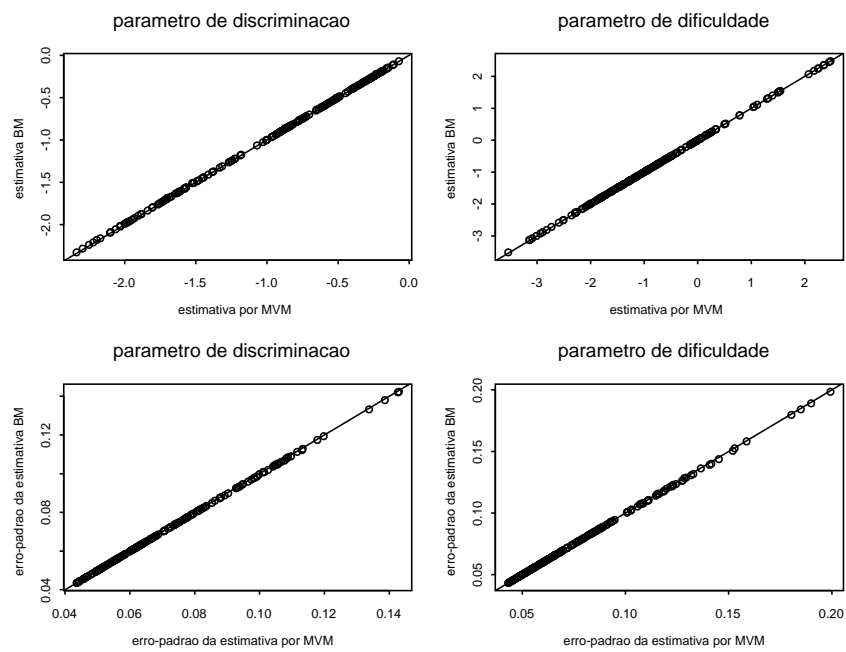


Figura 9.2 *Comparação entre as estimativas MVM e BM - parâmetros irrestritos*

### 9.3.2 Modelos Dicotômicos versus Modelo de Resposta Nominal

Como explicitado no Capítulo 6, um dos principais objetivos do Modelo de Resposta Nominal é o de extrair mais informação do teste a fim de aperfeiçoar as estimativas das habilidades dos indivíduos [Bock (1997)]. No entanto, em aplicações práticas, de um modo geral, utilizam-se os modelos dicotômicos para a análise dos testes. Dessa forma, desconsideram-se informações contidas nas alternativas incorretas (conhecidas como **distratores**), que poderiam aperfeiçoar as estimativas dos traços latentes. É importante esclarecer que nem sempre o Modelo de Resposta Nominal seria o mais razoável a ser usado para analisar testes de múltipla escolha. Porém, estamos considerando que o teste de múltipla escolha em estudo pode ser razoavelmente modelado por ele.

Nesse sentido, conduzimos um estudo de simulação com o intuito de comparar as estimativas das habilidades através dos dois modelos citados. As respostas foram geradas de acordo com o descrito na Seção 9.3.1, supondo-se que este mecanismo gere respostas mais próximas daquelas que seriam geradas por respondentes quando submetidos a testes de múltipla escolha. O conjunto de dados, incluindo as estimativas dos parâmetros, é exatamente aquele estudado na referida seção. Posteriormente, estes foram corrigidos de maneira dicotômica, a fim de viabilizar o uso do MLUP3, considerando-se que, as categorias com os maiores valores dos parâmetros ( $a$  e  $b$ ), constituem as alternativas corretas. Ressaltamos ainda que, para ambos os modelos, estimamos as habilidades posteriormente à estimação dos parâmetros dos itens. Para a estimação referente ao Modelo de Resposta Nominal utilizou-se programas construídos via linguagem Ox versão 3.0 [Doornik (2000)]. No que diz respeito ao Modelo de Resposta Dicotômica utilizou-se o programa Bilog [Mislevy & Bock (1990)]. As estimativas das habilidades obedecem à seguinte nomenclatura

- EMVMVM : estimativa de máxima verossimilhança usando as estimativas de MVM dos parâmetros dos itens.
- EMVMMAP : estimativa de máxima verossimilhança usando as estimativas de MMAP dos parâmetros dos itens.
- EEAPMVM : estimativa esperança a posteriori usando as estimativas de MVM dos parâmetros dos itens.
- EEAPMMAP : estimativa esperança a posteriori usando as estimativas de MMAP dos parâmetros dos itens.
- EMAPMVM : estimativa moda a posteriori usando as estimativas de MVM dos parâmetros dos itens.
- EMAPMMAP : estimativa moda a posteriori usando as estimativas de MMAP dos parâmetros dos itens.

A Figura 9.3 apresenta as habilidades geradas. Notamos a pouca quantidade de indivíduos com habilidades próximas aos extremos. Pela Tabela 9.2, percebemos uma considerável superioridade dos métodos bayesianos com relação aos métodos de máxima verossimilhança. Em muitos casos, houve problemas na convergência dos procedimentos clássicos, como era de se esperar. Notamos também que o Modelo de Resposta Nominal fornece, não só melhores estimativas pontuais, mas principalmente, erros-padrão menores. Isso significa que para uma mesmo número de itens, temos estimativas mais precisas usando-se o MRN. O que está de acordo com o esperado. Em geral, as estatísticas apresentaram-se 3 vezes menor para o MRN do que para o MLUP3. Além disso, vemos que as estimativas MAP forneceram melhores resultados, pontuais e vieses, que as demais.

A Tabela 9.3 apresenta medidas resumo relacionadas aos erros-padrão das estimativas das habilidades. Notamos um claro desempenho superior dos métodos bayesianos bem como do Modelo de Resposta Nominal. Para estes os erros-padrão apresentam-se, em média, menores e com um menor intervalo de variação. Além disso podemos notar que alguns erros-padrão relativos as estimativas de máxima verossimilhança via MLUP3 apresentaram elevada magnitude.

A Figura 9.4 ilustra o melhor desempenho obtido via MRN. Notamos que a distribuição das estimativas das habilidades obtidas via MRN estão bem mais próximas dos verdadeiros valores do que aquelas obtidas via MLUP3. As Figuras 9.5 e 9.6, ilustram a superioridade das estimativas pontuais dos métodos bayesianos.

É digno de nota que também foi utilizado o MULP2 e que os resultados não melhoraram muito em favor deste.

Tabela 9.2 *Estatísticas relativas às estimativas das habilidades*

Métodos de Estimação	Estatísticas		
	SQR	SQVar	SEQM
<b>MLUP3</b>			
EMVMVM	4412,523	182637171,986	182641584,509
EMVM MAP	5113,547	313376895,907	313382009,454
EEAPMVM	1697,619	1674,142	3371,762
EEAPM MAP	1676,059	1542,217	3218,276
EMAPMVM	1673,391	1653,332	3326,723
EMAPM MAP	1685,052	1520,068	3205,120
<b>MRN</b>			
EMVMVM	589,032	565,999	1155,031
EMVM MAP	591,740	568,752	1160,491
EEAPMVM	519,882	500,106	1019,988
EEAPM MAP	519,710	502,168	1021,878
EMAPMVM	520,211	495,936	1016,146
EMAPM MAP	519,939	497,963	1017,902

Tabela 9.3 *Estatísticas relativas aos erros-padrão das estimativas das habilidades*

Métodos de Estimação	Estatísticas					
	Média	Erro-Padrão	C.V.(%)	Mínimo	Mediana	Máximo
<b>MLUP3</b>						
EMVMVM	37,288	187,449	502,711	0,518	0,684	999,000
EMVMMMAP	63,478	242,169	381,502	0,392	0,610	999,000
EEAPMVM	0,575	0,062	10,766	0,464	0,569	0,743
EEAPMMAP	0,548	0,090	16,400	0,371	0,546	0,730
EMAPMVM	0,572	0,064	11,124	0,460	0,565	0,738
EMAPMMAP	0,541	0,107	19,788	0,365	0,526	0,805
<b>MRN</b>						
EMVMVM	0,335	0,027	7,962	0,317	0,325	0,579
EMVMMAP	0,336	0,027	8,020	0,318	0,326	0,583
EEAPMVM	0,316	0,016	5,216	0,303	0,310	0,426
EEAPMMAP	0,316	0,017	5,240	0,304	0,310	0,427
EMAPMVM	0,315	0,017	5,250	0,302	0,308	0,444
EMAPMMAP	0,315	0,017	5,276	0,303	0,309	0,446

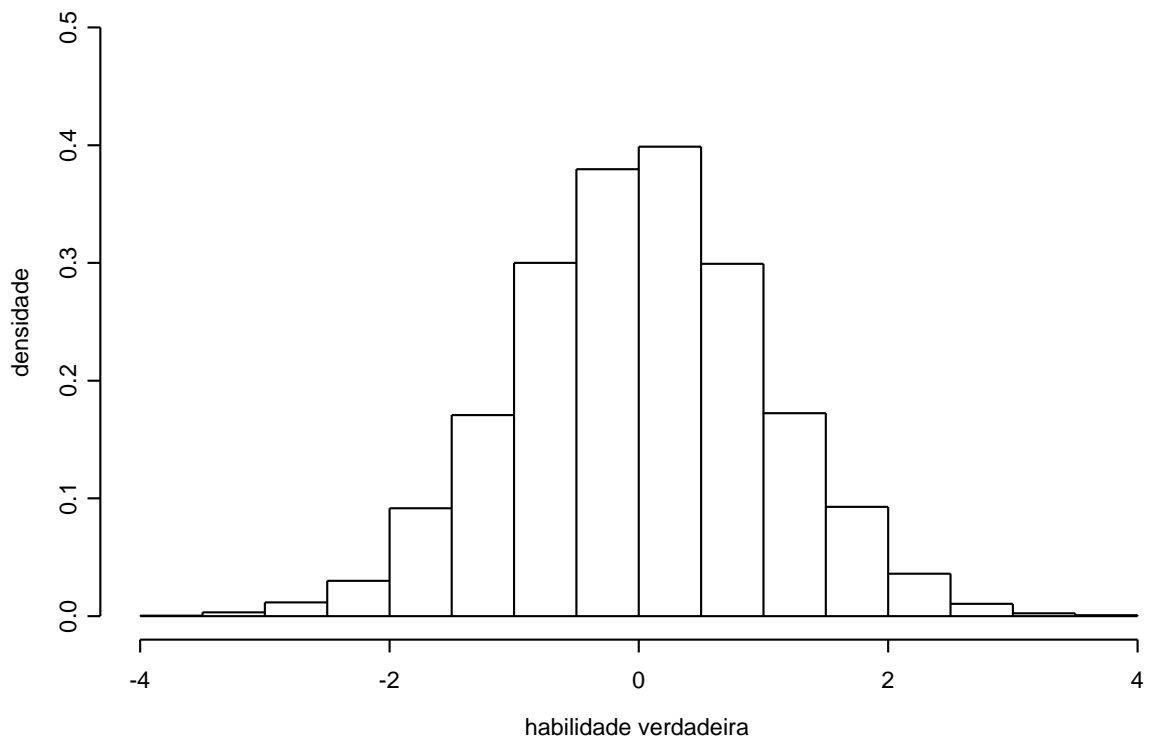


Figura 9.3 *Histograma das habilidades geradas*

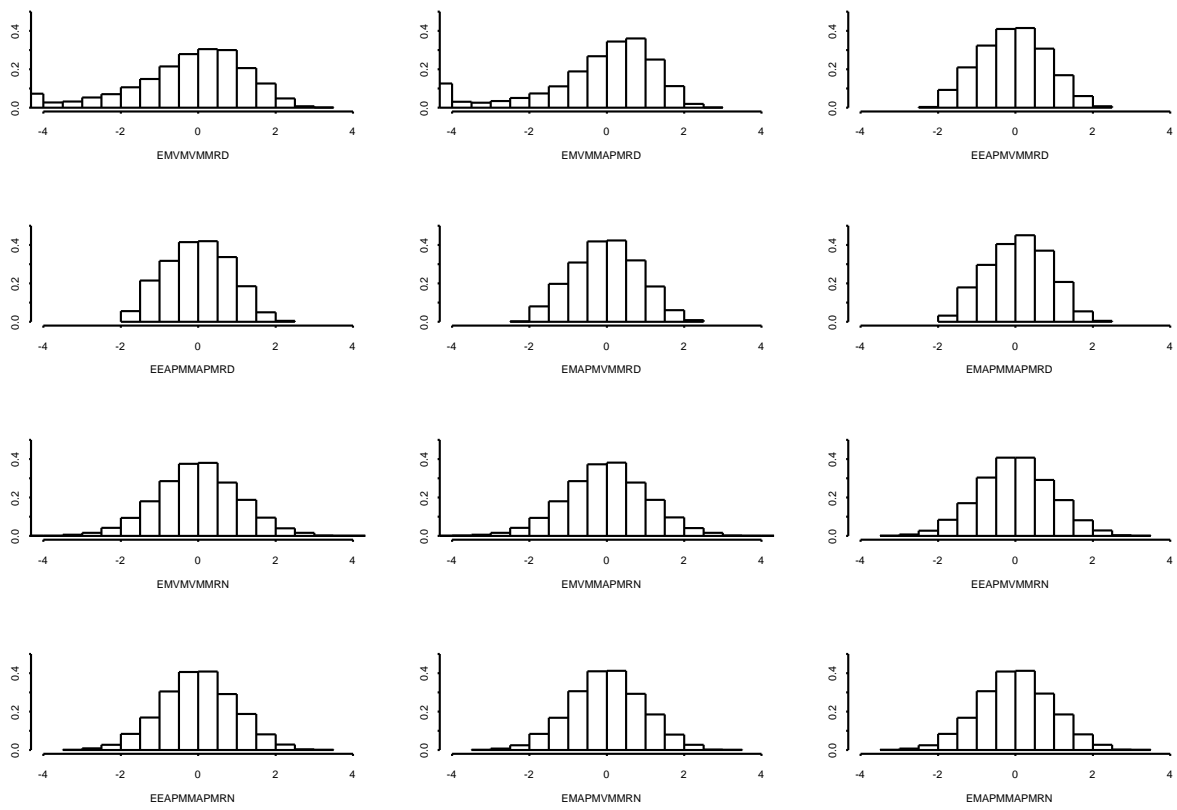


Figura 9.4 *Histogramas das estimativas das habilidades*



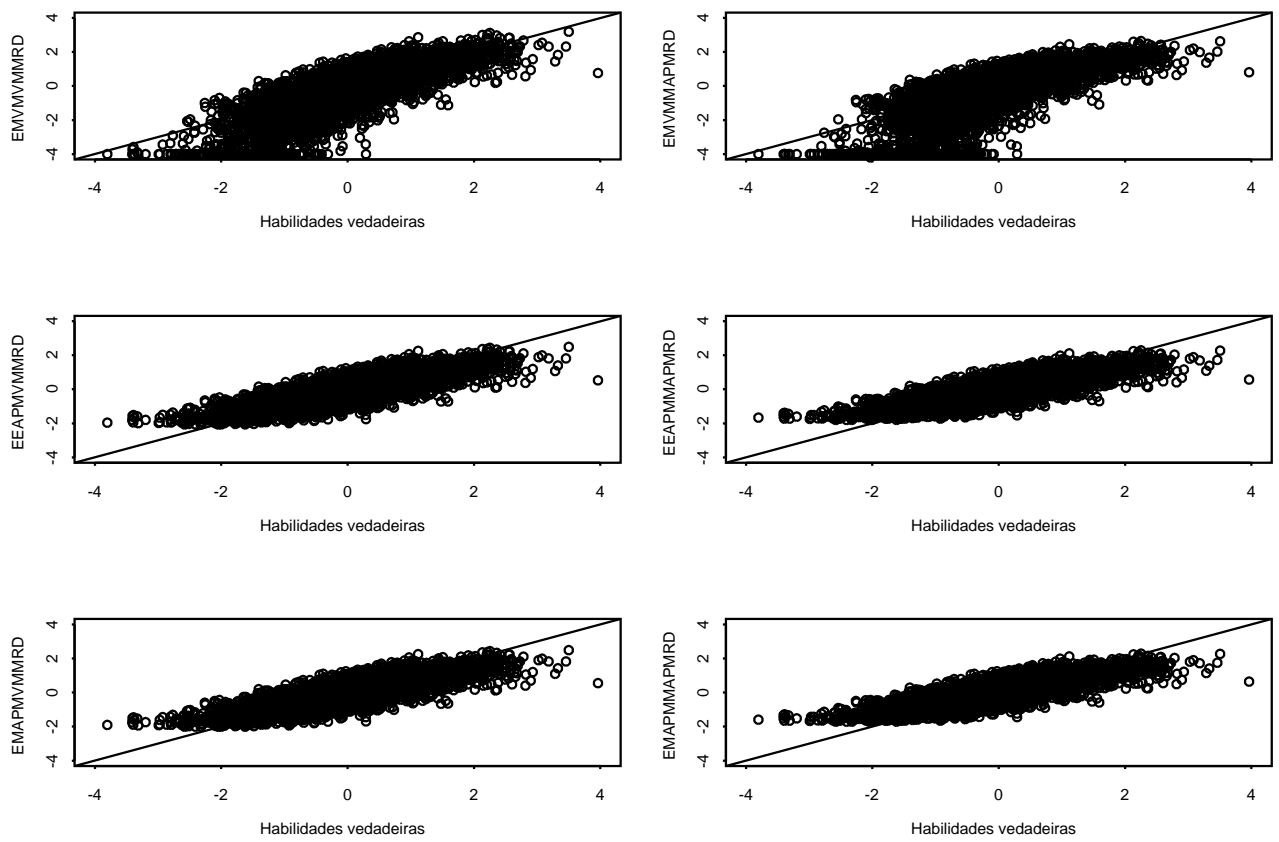


Figura 9.5 *Dispersão entre as verdadeiras habilidades e as estimativas - MLUP3*

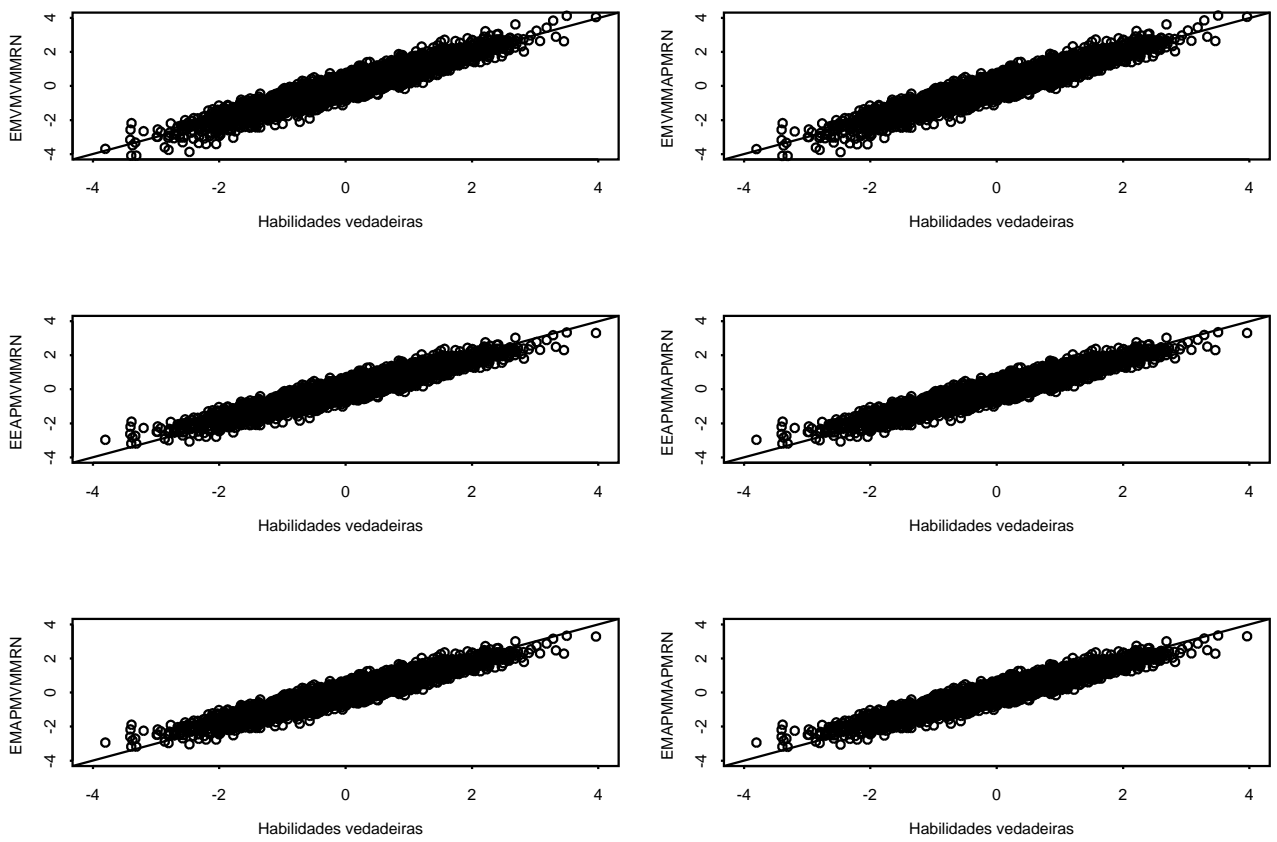


Figura 9.6 *Dispersão entre as verdadeiras habilidades e as estimativas - MRN*

---

## Capítulo 10

# Comentários e Sugestões

---

Como já foi bastante salientado neste trabalho, os métodos de estimação na TRI merecem uma atenção especial, pois a aplicação dela depende fundamentalmente dos progressos nessa área. Neste Capítulo enumeramos alguns comentários e sugestões.

### 10.1 Métodos de Monte Carlo via Cadeias de Markov

Os Métodos de MCMC merecem uma pesquisa mais abrangente e profunda, haja vista que eles fornecem soluções bastante razoáveis a diversas situações práticas tais como dados faltantes e modelos policotômicos. Para outros modelos como os Multivariados [Matos (2001)], Longitudinais [Tavares (2001)] e Multidimensionais [Nojosa (2001)], seria interessante o estudo e implementação desses métodos, pois os mesmos também se apresentam como alternativas para pequenas amostras. Sugerimos a leitura de Albert (1992), Bolt & Lall (2001), Ghosh et al (19998), Wollack et al (2000), Jones & Nediak (2000), Sahu (2001), além daquelas citadas no texto.

### 10.2 Distribuição das habilidades

Métodos que comportem outras distribuições para as habilidades, além da distribuição normal, são de fundamental importância. Notadamente, distribuições assimétricas são de interesse prático, principalmente em situações em que indivíduos foram pré-selecionados em testes anteriores. Os métodos de MCMC citados anteriormente, podem se constituir em uma boa alternativa.

### 10.3 Recursos Computacionais

Muito embora os programas desenvolvidos para este trabalho não tenham sido utilizados em sua plenitude, pretendemos aperfeiçoá-los para que no futuro, possamos desenvolver pacotes de análise comerciais.

## 10.4 Estimação das habilidades

Um ponto interessante seria realizar um estudo de simulação a respeito da estimação das habilidades no Modelo de Resposta Gradual versus os Modelos Dicotômicos, à semelhança do que foi feito neste trabalho. Isso serviria para verificar uma possível perda na estimação das habilidades em decorrência do uso destes últimos modelos em testes com itens corrigidos de forma gradual, tal como itens dissertativos corrigidos segundo alguma escala.

## 10.5 Equalização

Uma das grandes contribuições da TRI foi a de possibilitar que indivíduos submetidos a diferentes provas tivessem seus resultados comparáveis [Andrade(2001), Valle (1999)]. Existem muitos **Métodos de Equalização** (processo de inserir os resultados em uma mesma métrica) [Mislevy (1992) e Kolen & Brennan (1995)]. Contudo, o desenvolvimento do **Modelo de Grupos Múltiplos** [Bock & Zimowski (1997)] representou um avanço nesta área por permitir que este processo seja feito em concomitância com a estimação dos parâmetros. Muito embora se tenha desenvolvido os processos de estimação para este modelo, não se apresentou neste trabalho por não contemplar os objetivos principais. Entretanto, muito ainda precisa ser feito com relação à estimação neste modelo. Principalmente no que diz respeito aos métodos bayesianos e aos recursos computacionais.

## 10.6 Teoria Assintótica

Muito embora se tenha feito bastante com respeito a este tópico, existem lacunas a serem preenchidas. O estudo da convergência nos estimadores no contexto da TRI, onde muitas vezes não temos certas condições satisfeitas, como variáveis i.i.d., é de suma importância nos processos de inferência. Referências importantes nesse sentido são Houtink & Boomsma (1996), Chang & Stout (1993) e Chang (1996).

## 10.7 Estimação dos hiperparâmetros

Como explicitado na Seção 3.4 uma alternativa para se aprimorar as estimativas dos parâmetros dos itens seria estimar os **hiperparâmetros** em cada passo do processo iterativo. Pouco se fez a respeito, principalmente em termos de disponibilidade computacional. Seria interessante estudos mais profundos para se investigar o ganho que esse processo traz.

## 10.8 Elaboração de Textos em Português

O presente trabalho, na realidade, constitui um embrião para um futuro texto sobre estimação paramétrica que pretendemos escrever. A escassez de textos nesta área, e, principalmente, a ausência de literatura em língua portuguesa, nos incentiva a dar continuidade abordando, também, outros modelos.

## 10.9 Modelos Policotômicos

Os valores obtidos para as habilidades (proficiências) são arbitrários e necessitam de uma interpretação no contexto da área em que eles foram obtidos. Esta interpretação depende basicamente dos itens utilizados na elaboração dos instrumentos de medida. O INEP/MEC, por exemplo, construiu a Escala Nacional de Proficiência para interpretar os resultados do SAEB em cada uma das disciplinas estudadas, utilizando modelos dicotômicos. A construção de escalas de habilidade a partir de modelos policotômicos no Brasil ainda não foi feita e mereceria uma atenção especial.

Um outro tópico de pesquisa seria a utilização de modelos policotômicos para o estudo/definição do número apropriado de categorias nos itens. Por exemplo, vários instrumentos de medida fazem uso da escala de Likert com 5 pontos, sendo que o uso de uma categoria intermediária tem sido criticada na literatura [Garland (1991)].

---

## Apêndice A

# Expressões dos Processos de Estimação

---

### A.1 Expressões do Capítulo 2

#### A.1.1 Parâmetros dos Itens

Para o cálculo da Matriz Hessiana (Informação de Fisher) precisaremos do seguinte resultado

$$\frac{\partial^2 P_{ij}}{\partial \zeta_i \partial \zeta_i'},$$

por outro lado, temos que

$$\begin{aligned} \frac{\partial P_{ij}^* Q_{ij}^*}{\partial \zeta_i} &= \frac{\partial P_{ij}^*}{\partial \zeta_i} Q_{ij}^* + \frac{\partial Q_{ij}^*}{\partial \zeta_i} P_{ij}^* \\ &= \frac{\partial P_{ij}^*}{\partial \zeta_i} \{1 - P_{ij}^*\} - \frac{\partial P_{ij}^*}{\partial \zeta_i} P_{ij}^* \\ &= \{1 - 2P_{ij}^*\} \frac{\partial P_{ij}^*}{\partial \zeta_i}. \end{aligned} \tag{A.1}$$

Dessa forma, de (A.1), temos que

$$\frac{\partial^2 P_{ij}}{\partial a_i^2} = D^2 (1 - c_i) (\theta_j - b_i)^2 (1 - 2P_{ij}^*) P_{ij}^* Q_{ij}^*, \tag{A.2}$$

$$\frac{\partial^2 P_{ij}}{\partial b_i^2} = D^2 a_i^2 (1 - c_i) (1 - 2P_{ij}^*) P_{ij}^* Q_{ij}^*, \tag{A.3}$$

$$\frac{\partial^2 P_{ij}}{\partial c_i^2} = 0, \tag{A.4}$$

$$\frac{\partial^2 P_{ij}}{\partial a_i \partial b_i} = -D(1 - c_i) \{1 + Da_i (\theta_j - b_i) (1 - 2P_{ij}^*)\} P_{ij}^* Q_{ij}^*, \tag{A.5}$$

$$\frac{\partial^2 P_{ij}}{\partial a_i \partial c_i} = -D (\theta_j - b_i) P_{ij}^* Q_{ij}^*, \tag{A.6}$$

$$\frac{\partial^2 P_{ij}}{\partial b_i \partial c_i} = Da_i P_{ij}^* Q_{ij}^*. \tag{A.7}$$

Por outro lado, temos que a matriz Hessiana é dada por

$$H(\zeta_i) = \sum_{j=1}^n \left\{ \left[ \frac{\partial v_{ij}}{\partial \zeta_i} \right] \left( \frac{\partial P_{ij}}{\partial \zeta_i} \right)' + v_{ij} \left( \frac{\partial^2 P_{ij}}{\partial \zeta_i \partial \zeta_i'} \right) \right\}. \quad (\text{A.8})$$

Dessa forma, de (A.2) a (A.7) em (A.8), temos (2.12). A informação de Fisher segue desse resultado.

### A.1.2 Habilidades

Para calcular a matriz Hessiana das habilidades, notemos que

$$\begin{aligned} H(\theta_j) &= \frac{\partial S(\theta_j)}{\partial \theta_j} = \frac{\partial P_{ij}}{\partial \theta_j} \left\{ \sum_{i=1}^I \left\{ \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right\} \left( \frac{\partial P_{ij}}{\partial \theta_j} \right) \right\} \\ &= \frac{\partial}{\partial \theta_j} \left\{ \sum_{i=1}^I v_{ij} \left( \frac{\partial P_{ij}}{\partial \theta_j} \right) \right\} \\ &= \sum_{i=1}^I \left\{ \left( \frac{\partial v_{ij}}{\partial \theta_j} \right) \left( \frac{\partial P_{ij}}{\partial \theta_j} \right) + v_{ij} \left( \frac{\partial^2 P_{ij}}{\partial \theta_j^2} \right) \right\}. \end{aligned} \quad (\text{A.9})$$

Com  $v_{ij} = \left\{ \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right\}$ .

Por outro lado,

$$\frac{\partial v_{ij}}{\partial \theta_j} = -v_{ij}^2 \left( \frac{\partial P_{ij}}{\partial \theta_j} \right), \quad (\text{A.10})$$

Assim de (A.1), temos que

$$\frac{\partial^2 P_{ij}}{\partial \theta_j^2} = D^2 a_i^2 (1 - c_i) \{1 - 2P_{ij}^*\} P_{ij}^* Q_{ij}^* = H_{ij} P_{ij}^* Q_{ij}^*, \quad (\text{A.11})$$

onde  $v_{ij} = D^2 a_i^2 (1 - c_i) \{1 - 2P_{ij}^*\}$ .

Portanto, de (A.10) e (A.11) temos que (A.9) é dada por

$$H(\theta_j) = \sum_{i=1}^I (y_{ij} - P_{ij}) W_{ij} \{H_{ij} - (y_{ij} - P_{ij}) W_{ij} h_{ij}^2\}, \quad (\text{A.12})$$

e a Informação de Fisher

$$I(\theta_j) = \mathbf{E} \{-H(\theta_j)\} = \sum_{i=1}^I P_{ij}^* Q_{ij}^* W_{ij} h_{ij}^2. \quad (\text{A.13})$$

## A.2 Expressões do Modelos de Resposta Nominal

### A.2.1 Parâmetros dos itens - Habilidades Conhecidas

Desenvolveremos nesta parte as expressões relativas ao vetor escore, matriz Hessiana (informação de Fisher).

Primeiramente precisaremos de alguns desenvolvimentos intermediários. Para chegarmos a (4.9), notemos que de (4.3), temos

$$\begin{aligned}
\mathbf{Z}_{ij} &= \begin{bmatrix} 1 & \theta_j \end{bmatrix} \begin{bmatrix} \delta_{i1} & \dots & \delta_{im_{i-1}} \\ \alpha_{i1} & \dots & \alpha_{im_{i-1}} \end{bmatrix} \begin{bmatrix} \frac{1}{m_i} & \frac{1}{m_i} - 1 & \dots & \frac{1}{m_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m_i} & \frac{1}{m_i} & \dots & \frac{1}{m_i} - 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \theta_j \end{bmatrix} \begin{bmatrix} \sum_{s=1}^{m_i-1} \frac{\delta_{is}}{m_i} & -\delta_{i1} + \sum_{s=1}^{m_i-1} \frac{\delta_{is}}{m_i} & \dots & -\delta_{im_i} + \sum_{s=1}^{m_i} \frac{\delta_{is}}{m_i} \\ \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} & -\alpha_{i1} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} & \dots & -\alpha_{im_i} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{s=1}^{m_i-1} \frac{\delta_{is}}{m_i} + \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) \theta_j & \delta_{i1} + \sum_{s=1}^{m_i-1} \frac{\delta_{is}}{m_i} + \left( -\alpha_{i1} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) \theta_j & \dots \\ \delta_{i(m_i-1)} + \sum_{s=1}^{m_i-1} \frac{\delta_{is}}{m_i} + \left( -\alpha_{i(m_i-1)} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) \theta_j \end{bmatrix}.
\end{aligned}$$

O que nos leva a (4.9). Assim, diferenciando-se (4.9) com relação a  $\delta_{ih}$ ,  $h = 1, \dots, m_i - 1$ , vem que

$$\frac{\partial \mathbf{Z}_{ij}}{\partial \delta_i} = \begin{bmatrix} \frac{\partial \mathbf{Z}_{ij}}{\partial \delta_{i1}} \\ \vdots \\ \frac{\partial \mathbf{Z}_{ij}}{\partial \delta_{i(m_i-1)}} \end{bmatrix} = \begin{bmatrix} \frac{1}{m_i} & \frac{1}{m_i} - 1 & \frac{1}{m_i} & \dots & \frac{1}{m_i} \\ \frac{1}{m_i} & \frac{1}{m_i} & \frac{1}{m_i} - 1 & \dots & \frac{1}{m_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m_i} & \frac{1}{m_i} & \frac{1}{m_i} & \dots & \frac{1}{m_i} - 1 \end{bmatrix}.$$

O que demonstra (4.10) e analogamente (4.11).

Para chegar em (4.12) e (4.13), usaremos (4.10) e (4.11), o que nos leva a



$$\begin{aligned}
\frac{\partial w_{ij}}{\partial \delta_{ih}} &= \sum_{s \neq h+1} \left\{ e^{z_{ijs}} \frac{\partial z_{ijs}}{\partial \delta_{ih}} \right\} + \frac{\partial z_{ij(h+1)}}{\partial \delta_{ih}} e^{z_{ij(h+1)}} \\
&= \sum_{s \neq h+1} \frac{e^{z_{ijs}}}{m_i} + e^{z_{ij(h+1)}} \left( \frac{1}{m_i} - 1 \right) \\
&= \sum_{s=1}^{m_i} \frac{e^{z_{ijs}}}{m_i} - e^{z_{ij(h+1)}} = \frac{w_{ij}}{m_i} - e^{z_{ij(h+1)}}.
\end{aligned}$$

O que finda a prova. Para expressões (4.14) e (4.15), usando (4.12) e (4.13), temos, Para  $s \neq h + 1$ ,

$$\begin{aligned}
\frac{\partial P_{ijs}}{\partial \delta_{ih}} &= \frac{1}{w_{ij}^2} \left\{ e^{z_{ijs}} \left( \frac{\partial z_{ijs}}{\delta_{ih}} \right) w_{ij} - e^{z_{ijs}} \left( \frac{w_{ij}}{m_i} - e^{z_{ij(h+1)}} \right) \right\} \\
&= \frac{e^{z_{ijh}} e^{z_{ij(h+1)}}}{w_{ij}^2} = P_{ijs} P_{ij(h+1)}.
\end{aligned}$$

Para  $s = h + 1$ ,

$$\begin{aligned}
\frac{\partial P_{ij(h+1)}}{\partial \delta_{ih}} &= \frac{1}{w_{ij}^2} \left\{ e^{z_{ij(h+1)}} \left( \frac{\partial z_{ij(h+1)}}{\delta_{ih}} \right) w_{ij} - e^{z_{ij(h+1)}} \left( \frac{w_{ij}}{m_i} - e^{z_{ij(h+1)}} \right) \right\} \\
&= \frac{1}{w_{ij}^2} \left\{ -e^{z_{ij(h+1)}} w_{ij} + \frac{e^{z_{ij(h+1)}}}{m_i} w_{ij} - \frac{e^{z_{ij(h+1)}}}{m_i} w_{ij} + (e^{z_{ij(h+1)}})^2 \right\} \\
&= -P_{ij(h+1)} + P_{ij(h+1)}^2 = -P_{ij(h+1)} (1 - P_{ij(h+1)}).
\end{aligned}$$

**Q. E. D**

Para desenvolvermos a expressão relativa ao vetor Escore, usaremos (4.9) a (4.15), obtendo

$$\begin{aligned}
\frac{\partial l(\Gamma)}{\partial \delta_{ih}} &= \sum_{j=1}^n \left\{ \frac{\partial}{\partial \delta_{ih}} [y_{ij1} \ln P_{ij1} + \dots + y_{ij(h+1)} \ln P_{ij(h+1)} + \dots + y_{ijm_i}] \right\} \\
&= \sum_{j=1}^n \left\{ \frac{y_{ij1}}{P_{ij1}} (P_{ij1} P_{ij(h+1)}) + \dots + \frac{y_{ij(h+1)}}{P_{ij(h+1)}} [P_{ij(h+1)} (1 - P_{ij(h+1)})] \right. \\
&\quad \left. + \dots + \frac{y_{ijm_i}}{P_{ijm_i}} (P_{ijm_i} P_{ij(h+1)}) \right\} \\
&= \sum_{j=1}^n \{ P_{ij(h+1)} - y_{ij(h+1)} \}.
\end{aligned}$$

Finalmente, usando os resultados matriciais de Bock (1972), temos

$$\mathbf{S}(\boldsymbol{\delta}_i) = \sum_{j=1}^n \mathbf{T}_i \{ \mathbf{y}_{ij.} - \mathbf{P}_{ij.} \},$$

em que

$$\mathbf{P}_{ij.} = (P_{ij1}, \dots, P_{ijm_i})',$$

$$\mathbf{y}_{ij.} = (y_{ij1}, \dots, y_{ijm_i})'.$$

Analogamente para o  $\boldsymbol{\alpha}_i$ , temos

$$\mathbf{S}(\boldsymbol{\alpha}_i) = \sum_{j=1}^n \mathbf{T}_i \{ \mathbf{y}_{ij.} - \mathbf{P}_{ij.} \} \theta_j.$$

O que nos leva a

$$\mathbf{S}(\boldsymbol{\Gamma}_i) = \sum_{j=1}^n \left\{ \begin{bmatrix} 1 \\ \theta_j \end{bmatrix} \otimes \mathbf{T}_i [\mathbf{y}_{ij.} - \mathbf{P}_{ij.}] \right\}.$$

Novamente, usando as equações (4.9) a (4.15) e (4.22), temos que as componentes da matriz Hessiana são

$$H(\delta_{ih}, \delta_{ih}) = \sum_{j=1}^n \left\{ \frac{\partial P_{ij(h+1)}}{\partial \delta_{ih}} \right\} = - \sum_{j=1}^n \{ P_{ij(h+1)} (1 - P_{ij(h+1)}) \},$$

$$H(\delta_{ih}, \delta_{ih'}) = \sum_{j=1}^n \left\{ \frac{\partial P_{ij(h+1)}}{\partial \delta_{ih'}} \right\} = \sum_{j=1}^n \{ P_{ij(h+1)} P_{ij(h'+1)} \}.$$

Analogamente,

$$H(\alpha_{ih}, \alpha_{ih}) = \sum_{j=1}^n \left\{ \frac{\partial P_{ij(h+1)}}{\partial \alpha_{ih}} \right\} = - \sum_{j=1}^n \{ P_{ij(h+1)} (1 - P_{ij(h+1)}) \} \theta_j^2,$$

$$H(\alpha_{ih}, \alpha_{ih'}) = \sum_{j=1}^n \left\{ \frac{\partial P_{ij(h+1)}}{\partial \alpha_{ih'}} \right\} = \sum_{j=1}^n \{ P_{ij(h+1)} P_{ij(h'+1)} \} \theta_j^2,$$

$$H(\alpha_{ih}, \delta_{ih}) = \sum_{j=1}^n \left\{ \frac{\partial P_{ij(h+1)}}{\partial \delta_{ih}} \right\} \theta_j = - \sum_{j=1}^n \{ P_{ij(h+1)} (1 - P_{ij(h+1)}) \} \theta_j,$$

$$H(\alpha_{ih}, \delta_{ih'}) = \sum_{j=1}^n \left\{ \frac{\partial P_{ij(h+1)}}{\partial \delta_{ih'}} \right\} \theta_j = \sum_{j=1}^n \{ P_{ij(h+1)} P_{ij(h'+1)} \} \theta_j.$$

O Procedimento para a abordagem de **agrupamento das habilidades** é análogo, com as devidas modificações.

### A.2.2 Parâmetros dos itens - Habilidades Desconhecidas

Primeiramente, apresentaremos um procedimento geral que se aplica a qualquer modelo policotômico.

Temos que a log-verossimilhança marginal é dada por

$$\begin{aligned} l(\zeta, \eta) &= \sum_{j=1}^n \ln \int_{\mathcal{R}} P(\mathbf{Y}_{\cdot j} | \zeta, \theta) g(\theta | \eta) d\theta \\ &= \sum_{j=1}^n \ln P(\mathbf{Y}_{\cdot j} | \zeta, \eta). \end{aligned} \quad (\text{A.14})$$

Diferenciando-se (A.14) com relação aos **parâmetros dos itens**, temos

$$S(\zeta_i) = \sum_{j=1}^n \left\{ \frac{1}{P(\mathbf{Y}_{\cdot j} | \zeta, \eta)} \left( \frac{\partial}{\partial \zeta_i} P(\mathbf{Y}_{\cdot j} | \zeta, \eta) \right) \right\}$$

mas

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} P(\mathbf{Y}_{\cdot j} | \zeta, \eta) &= \frac{\partial}{\partial \zeta_i} \left\{ \int_{\mathcal{R}} P(\mathbf{Y}_{\cdot j} | \zeta, \theta) g(\theta | \eta) d\theta \right\} \\ &= \int_{\mathcal{R}} \frac{\partial}{\partial \zeta_i} \left\{ \prod_{i=1}^I P_{ij}^{y_{ij}} \right\} g(\theta | \eta) d\theta \end{aligned} \quad (\text{A.15})$$

com  $\prod_{i=1}^I P_{ij}^{y_{ij}} = \prod_{i=1}^I \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}$  e  $P_{ijs} = P(Y_{ijs} = 1 | \zeta_i, \theta)$ .

Seguindo temos,

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} P(\mathbf{Y}_{\cdot j} | \zeta, \eta) &= \int_{\mathcal{R}} \left\{ \prod_{i' \neq i} P_{ij}^{y_{ij}} \right\} \left\{ \frac{\partial P_{ij}^{y_{ij}}}{\partial \zeta_i} \right\} g(\theta | \eta) d\theta \\ &= \int_{\mathcal{R}} \left\{ \prod_{i=1}^I P_{ij}^{y_{ij}} \right\} \left\{ \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}} \right\}^{-1} \left\{ \frac{\partial P_{ij}^{y_{ij}}}{\partial \zeta_i} \right\} g(\theta | \eta) d\theta. \end{aligned} \quad (\text{A.16})$$

Por outro lado, sabemos que

$$\left\{ \prod_s y_s^{a_s} \right\}^{-1} \frac{\partial}{\partial \phi} \left\{ \prod_s y_s^{a_s} \right\} = \sum_s \left\{ \frac{\partial y_s}{\partial \phi} \right\} \left\{ \frac{a_s}{y_s} \right\}, \text{ com } y_s = f(\phi). \quad (\text{A.17})$$

Dessa forma, temos

$$\left\{ \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}} \right\}^{-1} \left\{ \frac{\partial}{\partial \zeta_i} \left[ \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}} \right] \right\} = \sum_{s=1}^{m_i} \left\{ \frac{\partial P_{ijs}}{\partial \zeta_i} \right\} \left\{ \frac{y_{ijs}}{P_{ijs}} \right\}. \quad (\text{A.18})$$

Assim, usando (A.18) em (A.16), temos

$$\frac{\partial}{\partial \zeta_i} P(\mathbf{Y}_{.j} | \zeta, \boldsymbol{\eta}) = \int_{\mathbb{R}} P(\mathbf{Y}_{.j} | \zeta, \theta) g(\theta | \boldsymbol{\eta}) \left\{ \sum_{s=1}^{m_i} \left[ \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial \zeta_i} \right) \right] \right\} d\theta.$$

Dessa modo, temos

$$\mathbf{S}(\zeta_i) = \sum_{j=1}^n \left\{ \int_{\mathbb{R}} g_j^*(\theta) \left\{ \sum_{s=1}^{m_i} \left[ \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial \zeta_i} \right) \right] \right\} d\theta \right\}. \quad (\text{A.19})$$

Em que  $g_j^*(\theta) = \frac{P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\eta}) g(\theta | \boldsymbol{\eta})}{\int_{\mathbb{R}} P(\mathbf{Y}_{.j} | \theta, \boldsymbol{\eta}) g(\theta | \boldsymbol{\eta})}$  o qual compõe os **dados artificiais**.

Para a matriz Hessiana, via abordagem de **Bock & Aitkin**, temos

$$\begin{aligned} H(\zeta_i) &= \sum_{j=1}^n \left\{ \int_{\mathbb{R}} g_j^*(\theta) \left\{ \sum_{j=1}^n \left[ \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial \zeta_i \partial \zeta_i'} \right) \right. \right. \right. \\ &\quad \left. \left. - y_{ijs} \left( \frac{\partial P_{ijs} / \partial \zeta_i}{P_{ijs}} \right) \left( \frac{\partial P_{ijs} / \partial \zeta_i}{P_{ijs}} \right)' \right] \right\} d\theta \right\} \\ &= \sum_{j=1}^n \left\{ \int_{\mathbb{R}} g_j^*(\theta) \left\{ \sum_{j=1}^n \frac{y_{ijs}}{P_{ijs}} \left[ \left( \frac{\partial P_{ijs}}{\partial \zeta_i \partial \zeta_i'} \right) \right. \right. \right. \\ &\quad \left. \left. - \frac{1}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial \zeta_i} \right) \left( \frac{\partial P_{ijs}}{\partial \zeta_i} \right)' \right] \right\} d\theta \right\}. \end{aligned} \quad (\text{A.20})$$

E dessa forma, as expressões, vetor escore e matriz Hessiana, podem ser desenvolvidas usando-se os resultados da Seção A.2.1.

## A.2.3 Estimação das Habilidades

Tal como no caso dos parâmetros dos itens, desenvolveremos alguns resultados intermediários, antes de nos atermos as expressões dos processos de estimação propriamente ditas.

Para chegarmos a (4.36), temos de (4.9), que

$$\frac{\partial \mathbf{Z}_{ij}}{\partial \theta_j} = \left[ \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} - \alpha_{i1} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \quad \dots \quad -\alpha_{i(m_i-1)} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right].$$

Para chegarmos a (4.37), usaremos (4.36), o que nos leva a

$$\begin{aligned} \frac{\partial w_{ij}}{\partial \theta_j} &= \sum_{s=1}^{m_i} \left\{ e^{z_{ijs}} \left( \frac{\partial e^{z_{ijs}}}{\partial \theta_j} \right) \right\} \\ &= e^{z_{ij1}} \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) + e^{z_{ij2}} \left( -\alpha_{i1} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) + \dots + e^{z_{ijm_i}} \left( -\alpha_{i(m_i-1)} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) \\ &= \{ e^{z_{ij1}} + \dots + e^{z_{ijm_i}} \} \left\{ \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right\} - (\alpha_{i1} e^{z_{ij2}} + \dots + \alpha_{i(m_i-1)} e^{z_{ijm_i}}) \\ &= \left\{ \sum_{s'=1}^{m_i} e^{z_{ijs}} \right\} \left\{ \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right\} - \sum_{s=1}^{m_i-1} e^{z_{ij(s+1)}} \alpha_{is}. \end{aligned}$$

**Q. E. D**

Com relação a (4.38), usando (4.36) e (4.37), temos

Para  $s = 1$

$$\begin{aligned} \frac{\partial P_{ij1}}{\partial \theta_j} &= \frac{1}{w_{ij}^2} \left\{ w_{ij} \left( \frac{\partial e^{z_{ij1}}}{\partial \theta_j} \right) - e^{z_{ij1}} \left( \frac{\partial w_{ij}}{\partial \theta_j} \right) \right\} \\ &= \frac{1}{w_{ij}^2} \left\{ w_{ij} e^{z_{ij1}} \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) - e^{z_{ij1}} \left\{ \left( \sum_{s=1}^{m_i} e^{z_{ijs}} \right) \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) \right. \right. \\ &\quad \left. \left. - \sum_{s=1}^{m_i-1} (e^{z_{ij(s+1)}} \alpha_{is}) \right\} \right\} \\ &= P_{ij1} \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) - P_{ij1} \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) + P_{ij1} \left\{ \sum_{s=1}^{m_i-1} \frac{e^{z_{ij(s+1)}} \alpha_{is}}{w_{ij}} \right\} \\ &= P_{ij1} \left\{ \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right\}. \end{aligned}$$

Para  $s = 2, \dots, m_i$

$$\begin{aligned}
\frac{\partial P_{ijs}}{\partial \theta_j} &= \frac{1}{w_{ij}^2} \left\{ w_{ij} \left( -\alpha_{is} + \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) e^{z_{ijs}} \right. \\
&\quad \left. - e^{z_{ijs}} \left\{ \left( \sum_{s=1}^{m_i} e^{z_{ijs}} \right) \left( \sum_{s=1}^{m_i-1} \frac{\alpha_{is}}{m_i} \right) - \sum_{s=1}^{m_i-1} e^{z_{ij(s+1)}} \alpha_{is} \right\} \right\} \\
&= P_{ijs} \left( -\alpha_{i(s-1)} + \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right).
\end{aligned}$$

O que finda a demonstração. Para chegarmos ao vetor escore (matriz Hessiana), equações (4.41) e (4.42), respectivamente, temos que usar (4.36) a (4.38) em (4.39), o que nos leva a

Se  $s = 1$

$$S(\theta_j) = \sum_{i=1}^I \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is}.$$

Para  $s = 2, \dots, m_i$

$$S(\theta_j) = \sum_{i=1}^I \left( -\alpha_{is} + \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right).$$

Usando os resultados matriciais de Baker (1992), temos que o vetor escore pode ser escrito como

$$S(\theta_j) = \sum_{i=1}^I \boldsymbol{\alpha}' \mathbf{T}'_i [\mathbf{y}_{ij.} - \mathbf{P}_{ij.}].$$

Para a matriz Hessiana, temos

$$\begin{aligned}
H(\theta_j) &= \sum_{i=1}^I \left\{ \sum_{s=1}^{m_i-1} \left( \frac{\partial P_{ij(s+1)}}{\partial \theta_j} \right) \right\} \\
&= \sum_{i=1}^I \left\{ \alpha_{i1} P_{ij2} \left( \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right) \right. \\
&\quad \left. + \dots + \alpha_{i(m_i-2)} P_{ij(m_i-1)} \left( -\alpha_{i(m_i-2)} + \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right) \right\} \\
&\quad \left. + \alpha_{i(m_i-1)} P_{ijm_i} \left( -\alpha_{i(m_i-1)} + \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right) \right\} \\
&= \sum_{i=1}^I \left\{ \left( \sum_{s=1}^{m_i-1} P_{ij(s+1)} \alpha_{is} \right)^2 - \sum_{s=1}^{m_i-1} \alpha_{is}^2 P_{ij(s+1)} \right\}.
\end{aligned}$$

Usando os resultados matriciais de Baker(1992), vem que

$$H(\theta_j) = - \sum_{i=1}^I \{ \alpha_i' T_i' W_{ij} T_i \alpha_i \} .$$

Q. E. D

### A.3 Expressões do Modelo de Resposta Gradual

#### A.3.1 Parâmetros dos itens - Habilidades Conhecidas

Antes de desenvolvermos as expressões dos processo iterativos, precisaremos dos seguintes resultados auxiliares. Considerando a expressão (6.1), temos

$$\begin{aligned} \frac{\partial P_{ij(s-1)}}{\partial b_{is}} &= \frac{\partial}{\partial b_{is}} \left\{ \frac{1}{1 + e^{-Da_i(\theta_j - b_{i(s-1)})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{is})}} \right\} \\ &= Da_i P_{ijs}^+ Q_{ijs}^+ , \end{aligned} \quad (\text{A.21})$$

$$\frac{\partial P_{ijs}}{\partial b_{is}} = -Da_i P_{ijs}^+ Q_{ijs}^+ , \quad (\text{A.22})$$

$$\frac{\partial P_{ijs}}{\partial a_i} = D \left\{ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right\} . \quad (\text{A.23})$$

Também, temos que

$$\frac{\partial W_{ijs}^+}{\partial \zeta_i} = \left\{ 1 - 2P_{ijs}^+ \right\} \left( \frac{\partial P_{ijs}^+}{\partial \zeta_i} \right) . \quad (\text{A.24})$$

Para as derivadas de segunda ordem, usando (A.24), vem que

$$\frac{\partial^2 P_{ij(s-1)}}{\partial b_{is}^2} = D^2 a_i^2 \left\{ 1 - 2P_{ijs}^+ \right\} W_{ijs}^+ \quad (\text{A.25})$$

e

$$\frac{\partial^2 P_{ij(s+h-1)}}{\partial b_{is} \partial b_{i(s+h)}} = 0 , \quad h \neq 0 , \quad (\text{A.26})$$

$$\begin{aligned} \frac{\partial^2 P_{ijs}}{\partial a_i^2} = & D^2 \left\{ (\theta_j - b_{is})^2 \left\{ 1 - 2P_{ijs}^+ \right\} W_{ijs}^+ \right. \\ & \left. - (\theta_j - b_{i(s+1)})^2 \left\{ 1 - 2P_{ij(s+1)}^+ \right\} W_{ij(s+1)}^+ \right\}, \end{aligned} \quad (\text{A.27})$$

$$\frac{\partial^2 P_{ij(s-1)}}{\partial a_i \partial b_{is}} = W_{ijs} \left\{ 1 + Da_i (\theta_j - b_{is}) \left\{ 1 - 2P_{ijs}^+ \right\} \right\}, \quad (\text{A.28})$$

$$\frac{\partial^2 P_{ijs}}{\partial a_i \partial b_{is}} = -W_{ijs} \left\{ 1 + Da_i (\theta_j - b_{is}) \left\{ 1 - 2P_{ijs}^+ \right\} \right\}. \quad (\text{A.29})$$

Para chegar ao vetor escore (equações (6.5) e (6.6)), usaremos (A.21) a (A.23) em (6.3). O que nos leva a

$$\begin{aligned} S(b_{is}) &= \sum_{j=1}^n \left\{ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} \left( \frac{\partial P_{ij(s-1)}}{\partial b_{is}} \right) + \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial b_{is}} \right) \right\} \\ &= Da_i \sum_{j=1}^n W_{ijs}^+ \left\{ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} - \frac{y_{ijs}}{P_{ijs}} \right\}, \end{aligned}$$

$$\begin{aligned} S(a_i) &= \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left( \frac{\partial P_{ijs}}{\partial a_i} \right) \right\} \\ &= D \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right] \right\}. \end{aligned}$$

Para a matriz Hessiana, usando-se (A.25) a (A.29) em (6.3), o que nos conduz a

$$\begin{aligned} H(a_i, a_i) &= D \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}^2} \left\{ (\theta_j - b_{is}) \left[ \left( \frac{\partial W_{ij}^+}{\partial a_i} \right) P_{ijs} - W_{ijs}^+ \left( \frac{\partial P_{ijs}}{\partial a_i} \right) \right] \right. \right. \\ &\quad \left. \left. - \frac{y_{ijs}}{P_{ijs}^2} \left\{ (\theta_j - b_{is}) \left[ \left( \frac{\partial W_{ij}^+}{\partial a_i} \right) P_{ijs} - W_{ijs}^+ \left( \frac{\partial P_{ijs}}{\partial a_i} \right) \right] \right\} \right\} \right. \\ &= \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{y_{ijs}}{P_{ijs}} \left[ \left[ D^2 \left\{ (\theta_j - b_{is})^2 \left\{ 1 - 2P_{ijs}^+ \right\} W_{ijs}^+ \right. \right. \right. \right. \\ &\quad \left. \left. \left. - (\theta_j - b_{i(s+1)})^2 \left\{ 1 - 2P_{ij(s+1)}^+ \right\} W_{ij(s+1)}^+ \right\} \right] P_{ijs} \right. \right. \\ &\quad \left. \left. - D \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right]^2 \right] \right\}, \end{aligned} \quad (\text{A.30})$$



$$\begin{aligned}
H(b_{is}, b_{is}) &= Da_i \sum_{j=1}^n \left\{ \frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} \left[ \left( \frac{\partial W_{ijs}^+}{\partial b_{is}} \right) P_{ij(s-1)} - W_{ijs}^+ \left( \frac{\partial P_{ij(s-1)}}{\partial b_{is}} \right) \right] \right. \\
&\quad \left. - \frac{y_{ijs}}{P_{ijs}^2} \left[ \left( \frac{\partial W_{ijs}^+}{\partial b_{is}} \right) P_{ijs} - W_{ijs}^+ \left( \frac{\partial P_{ijs}}{\partial b_{is}} \right) \right] \right\} \\
&= D \sum_{j=1}^n a_i \left\{ \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} - \frac{y_{ijs}}{P_{ijs}} \right] \left[ Da_i W_{ijs}^+ (1 - 2P_{ijs}^+) \right] \right. \\
&\quad \left. - \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} - \frac{y_{ijs}}{P_{ijs}^2} \right] Da_i (W_{ijs}^+)^2 \right\},
\end{aligned}$$

$$\begin{aligned}
H(b_{is}, b_{i(s-1)}) &= Da_i \sum_{j=1}^n W_{ijs}^+ \left\{ -\frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} \left( \frac{\partial P_{ij(s-1)}}{\partial b_{i(s-1)}} \right) + \frac{y_{ijs}}{P_{ijs}^2} \underbrace{\left( \frac{\partial P_{ijs}}{\partial b_{i(s-1)}} \right)}_0 \right\} \\
&= D^2 a_i^2 \sum_{j=1}^n \left\{ \frac{y_{ij(s-1)} W_{ijs}^+ W_{ij(s-1)}^+}{P_{ij(s-1)}^2} \right\}.
\end{aligned}$$

Analogamente,

$$H(b_{is}, b_{i(s+1)}) = D^2 a_i^2 \sum_{j=1}^n \left\{ \frac{y_{ijs} W_{ijs}^+ W_{ij(s+1)}^+}{P_{ijs}^2} \right\}.$$

Finalmente, temos

$$\begin{aligned}
H(a_i, b_{is}) &= \sum_{j=1}^n \left\{ \frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} \left[ \left( \frac{\partial P_{ij(s-1)}}{\partial a_i \partial b_{is}} \right) P_{ij(s-1)} - \left( \frac{\partial P_{ij(s-1)}}{\partial a_i} \right) \left( \frac{\partial P_{ij(s-1)}}{\partial b_{is}} \right) \right] \right. \\
&\quad \left. + \frac{y_{ijs}}{P_{ijs}^2} \left[ \left( \frac{\partial P_{ijs}}{\partial a_i \partial b_{is}} \right) P_{ijs} - \left( \frac{\partial P_{ijs}}{\partial a_i} \right) \left( \frac{\partial P_{ijs}}{\partial b_{is}} \right) \right] \right\} \\
&= \sum_{j=1}^n \left\{ D \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}} - \frac{y_{ijs}}{P_{ijs}} \right] W_{ijs}^+ \left\{ 1 + Da_i (\theta_j - b_{is}) (1 - 2P_{ijs}^+) \right\} \right. \\
&\quad - D^2 a_i W_{ijs}^+ \left[ \frac{y_{ij(s-1)}}{P_{ij(s-1)}^2} \left[ (\theta_j - b_{i(s-1)}) W_{ij(s-1)}^+ - (\theta_j - b_{is}) W_{ijs}^+ \right] \right. \\
&\quad \left. \left. - \frac{y_{ijs}}{P_{ijs}^2} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right] \right] \right\}. \tag{A.31}
\end{aligned}$$

Com relação a informação de Fisher, lembremos que  $Y_{ijs} \sim \text{Bernoulli}(P_{ijs})$ , onde  $(P_{ijs})$  é como dado em (6.1).

Dessa forma, temos

$$\begin{aligned}
 I(a_i, a_i) &= -D^2 \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ (\theta_j - b_{is})^2 (1 - 2P_{ijs}^+) W_{ijs}^+ - (\theta_j - b_{i(s+1)})^2 (1 - 2P_{ij(s+1)}^+) W_{ij(s+1)}^+ \right. \\
 &\quad \left. + D^2 \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{1}{P_{ijs}^2} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right]^2 \right\} \right\} \\
 &= D^2 \sum_{j=1}^n \sum_{s=0}^{m_i} \left\{ \frac{1}{P_{ijs}^2} \left[ (\theta_j - b_{is}) W_{ijs}^+ - (\theta_j - b_{i(s+1)}) W_{ij(s+1)}^+ \right]^2 \right\}.
 \end{aligned}$$

Pois o primeiro termo representa uma soma telescópica.

Para os outros segue-se que

$$\begin{aligned}
 I(b_{ih}, b_{ih}) &= -D^2 a_i^2 \sum_{j=1}^n \left\{ \underbrace{[1 - 1]}_0 \left[ (1 - 2P_{ijh}^+) W_{ijh}^+ \right] \right. \\
 &\quad \left. - \left[ \frac{1}{P_{ij(h-1)}} + \frac{1}{P_{ijh}} \right] (W_{ijh}^+)^2 \right\} \\
 &= D^2 a_i^2 \sum_{j=1}^n \left[ \frac{1}{P_{ij(h-1)}} + \frac{1}{P_{ijh}} \right] (W_{ijh}^+)^2. \tag{A.32}
 \end{aligned}$$

Para as expressões restantes o procedimento é análogo.

### A.3.2 Parâmetros dos Itens - Habilidades desconhecidos

De conformidade com o explicitado na Seção tal, o processo descrito se aplica a qualquer modelo policotômico, sendo necessário portanto, algumas poucas adaptações devido a mudança do modelos. Estas são tão somente, usar as expressões definidas anteriormente.

### A.3.3 Estimação das Habilidades

Para chegarmos as expressões pertinentes do processo de estimação, necessitamos antes explicitar algumas expressões, quais sejam

$$\frac{\partial P_{ijs}}{\partial \theta_j} = D a_i (W_{ijs}^+ - W_{ijs}^+), \tag{A.33}$$

$$\frac{\partial^2 P_{ijs}}{\partial \theta_j^2} = D^2 a_i^2 \left\{ (1 - 2P_{ijs}^+) W_{ijs}^+ - (1 - 2P_{ij(s+1)}^+) W_{ij(s+1)}^+ \right\}. \tag{A.34}$$

Dessa forma, usando-se (A.33) e (A.34), temos que o vetor escore e a matriz Hessiana, são dados respectivamente por

$$S(\theta_j) = D \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ \frac{a_i y_{ijs} (W_{ijs}^+ - W_{ij(s+1)}^+)}{P_{ijs}} \right\},$$

$$H(\theta_j) = D \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ \frac{a_i y_{ijs}}{P_{ijs}^2} \left[ \frac{\partial}{\partial \theta_j} (W_{ijs}^+ - W_{ij(s+1)}^+) P_{ijs} - \left( \frac{\partial P_{ijs}}{\partial \theta_j} \right) (W_{ijs}^+ - W_{ij(s+1)}^+) \right] \right\}$$

$$= D^2 \sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ a_i^2 y_{ijs} \left[ \frac{(1 - 2P_{ijs}^+) W_{ijs}^+ - (1 - 2P_{ij(s+1)}^+) W_{ij(s+1)}^+}{P_{ijs}} \right. \right.$$

$$\left. \left. - \frac{(W_{ijs}^+ - W_{ij(s+1)}^+)^2}{P_{ijs}^2} \right] \right\}.$$

Com relação a Informação de Fisher, temos

$$I(\theta_j) = -D^2 \underbrace{\sum_{i=1}^I \sum_{s=0}^{m_i} \left\{ a_i^2 \left[ (1 - 2P_{ijs}^+) W_{ijs}^+ \right] - \left[ (1 - 2P_{ij(s+1)}^+) W_{ij(s+1)}^+ \right] \right\}}_0$$

$$+ D^2 \sum_{i=1}^I \sum_{s=0}^{m_i} a_i^2 \left\{ \frac{(W_{ijs}^+ - W_{ij(s+1)}^+)^2}{P_{ijs}} \right\}$$

$$= D^2 \sum_{i=1}^I \sum_{s=0}^{m_i} a_i^2 \left\{ \frac{(W_{ijs}^+ - W_{ij(s+1)}^+)^2}{P_{ijs}} \right\}.$$

Pois o primeiro termo representa uma série telescópica.

Isto finda as demonstrações pertinentes.

---

## Apêndice B

### Simulação

---

A seguir apresentaremos os parâmetros usados nas simulações.

Tabela B.1 *Parâmetros originais usados nas simulações*

Item	alternativa	Disc. (a)	Dif. (b)	Item	alternativa	Disc. (a)	Dif. (b)
1	1	-0,75	-3,00	6	1	-0,75	-3,00
1	2	-0,50	-2,00	6	2	-0,50	-2,00
1	3	-0,20	-1,00	6	3	-0,20	-1,00
1	4	0,50	0,00	6	4	0,50	0,00
1	5	0,95	1,00	6	5	0,95	1,00
2	1	-0,75	-2,00	7	1	-0,75	-2,00
2	2	-0,50	-1,00	7	2	-0,50	-1,00
2	3	-0,20	0,00	7	3	-0,20	0,00
2	4	0,50	1,00	7	4	0,50	1,00
2	5	0,95	2,00	7	5	0,95	2,00
3	1	-0,75	-1,00	8	1	-0,70	-1,00
3	2	-0,50	0,00	8	2	-0,50	0,00
3	3	-0,20	1,00	8	3	0,20	1,00
3	4	0,50	2,00	8	4	0,30	2,00
3	5	0,95	3,00	8	5	0,70	3,00
4	1	-0,75	-2,50	9	1	-0,70	-2,50
4	2	-0,50	-1,50	9	2	-0,50	-1,50
4	3	-0,20	-0,50	9	3	0,20	-0,50
4	4	0,50	1,20	9	4	0,30	1,20
4	5	0,95	2,50	9	5	0,70	2,50
5	1	-0,75	-1,50	10	1	-0,70	-1,50
5	2	-0,50	-0,50	10	2	-0,50	-0,50
5	3	-0,20	1,50	10	3	0,20	1,50
5	4	0,50	2,50	10	4	0,30	2,50
5	5	0,95	3,50	10	5	0,70	3,50

---

Tabela B.1 Cont. .

Item	alternativa	Disc. (a)	Dif. (b)	Item	alternativa	Disc. (a)	Dif. (b)
11	1	-0,70	-3,00	18	1	-1,00	-1,00
11	2	-0,50	-2,00	18	2	-0,50	0,00
11	3	0,20	-1,00	18	3	-0,20	1,00
11	4	0,30	0,00	18	4	0,60	2,00
11	5	0,70	1,00	18	5	1,10	3,00
12	1	-0,70	-2,00	19	1	-1,00	-2,50
12	2	-0,50	-1,00	19	2	-0,50	-1,50
12	3	0,20	0,00	19	3	-0,20	-0,50
12	4	0,30	1,00	19	4	0,60	1,20
12	5	0,70	2,00	19	5	1,10	2,50
13	1	-0,70	-1,00	20	1	-1,00	-1,50
13	2	-0,50	0,00	20	2	-0,50	-0,50
13	3	0,20	1,00	20	3	-0,20	1,50
13	4	0,30	2,00	20	4	0,60	2,50
13	5	0,70	3,00	20	5	1,10	3,50
14	1	-0,70	-2,50	21	1	-1,00	-3,00
14	2	-0,50	-1,50	21	2	-0,50	-2,00
14	3	0,20	-0,50	21	3	-0,20	-1,00
14	4	0,30	1,20	21	4	0,60	0,00
14	5	0,70	2,50	21	5	1,10	1,00
15	1	-1,00	-1,50	22	1	-0,40	-2,00
15	2	-0,50	-0,50	22	2	-0,20	-1,00
15	3	-0,20	1,50	22	3	-0,10	0,00
15	4	0,60	2,50	22	4	0,20	1,00
15	5	1,10	3,50	22	5	0,50	2,00
16	1	-1,00	-3,00	23	1	-0,40	-1,00
16	2	-0,50	-2,00	23	2	-0,20	0,00
16	3	-0,20	-1,00	23	3	-0,10	1,00
16	4	0,60	0,00	23	4	0,20	2,00
16	5	1,10	1,00	23	5	0,50	3,00

Tabela B.1 Cont. .

Item	alternativa	Disc. (a)	Dif. (b)	Item	alternativa	Disc. (a)	Dif. (b)
17	1	-1,00	-2,00	24	1	-0,40	-2,50
17	2	-0,50	-1,00	24	2	-0,20	-1,50
17	3	-0,20	0,00	24	3	-0,10	-0,50
17	4	0,60	1,00	24	4	0,20	1,20
17	5	1,10	2,00	24	5	0,50	2,50
25	1	-0,40	-1,50	32	1	-1,00	-2,00
25	2	-0,20	-0,50	32	2	-0,60	-1,00
25	3	-0,10	1,50	32	3	-0,20	0,00
25	4	0,20	2,50	32	4	0,70	1,00
25	5	0,50	3,50	32	5	1,10	2,00
26	1	-0,40	-3,00	33	1	-1,00	-1,00
26	2	-0,20	-2,00	33	2	-0,60	0,00
26	3	-0,10	-1,00	33	3	-0,20	1,00
26	4	0,20	0,00	33	4	0,70	2,00
26	5	0,50	1,00	33	5	1,10	3,00
27	1	-0,40	-2,00	34	1	-1,00	-2,50
27	2	-0,20	-1,00	34	2	-0,60	-1,50
27	3	-0,10	0,00	34	3	-0,20	-0,50
27	4	0,20	1,00	34	4	0,70	1,20
27	5	0,50	2,00	34	5	1,10	2,50
28	1	-0,40	-1,00	35	1	-1,00	-1,50
28	2	-0,20	0,00	35	2	-0,60	-0,50
28	3	-0,10	1,00	35	3	-0,20	1,50
28	4	0,20	2,00	35	4	0,70	2,50
28	5	0,50	3,00	35	5	1,10	3,50
29	1	-1,00	-2,50	36	1	-0,50	-3,00
29	2	-0,60	-1,50	36	2	-0,20	-2,00
29	3	-0,20	-0,50	36	3	0,10	-1,00
29	4	0,70	1,20	36	4	0,20	0,00
29	5	1,10	2,50	36	5	0,40	1,00

Tabela B.1 Cont. .

Item	alternativa	Disc. (a)	Dif. (b)	Item	alternativa	Disc. (a)	Dif. (b)
30	1	-1,00	-1,50	37	1	-0,50	-2,00
30	2	-0,60	-0,50	37	2	-0,20	-1,00
30	3	-0,20	1,50	37	3	0,10	0,00
30	4	0,70	2,50	37	4	0,20	1,00
30	5	1,10	3,50	37	5	0,40	2,00
31	1	-1,00	-3,00	38	1	-0,50	-1,00
31	2	-0,60	-2,00	38	2	-0,20	0,00
31	3	-0,20	-1,00	38	3	0,10	1,00
31	4	0,70	0,00	38	4	0,20	2,00
31	5	1,10	1,00	38	5	0,40	3,00
39	1	-0,50	-2,50	45	1	-1,10	-1,50
39	2	-0,20	-1,50	45	2	-0,80	-0,50
39	3	0,10	-0,50	45	3	0,40	1,50
39	4	0,20	1,20	45	4	0,60	2,50
39	5	0,40	2,50	45	5	0,90	3,50
40	1	-0,50	-1,50	46	1	-1,10	-3,00
40	2	-0,20	-0,50	46	2	-0,80	-2,00
40	3	0,10	1,50	46	3	0,40	-1,00
40	4	0,20	2,50	46	4	0,60	0,00
40	5	0,40	3,50	46	5	0,90	1,00
41	1	-0,50	-3,00	47	1	-1,10	-2,00
41	2	-0,20	-2,00	47	2	-0,80	-1,00
41	3	0,10	-1,00	47	3	0,40	0,00
41	4	0,20	0,00	47	4	0,60	1,00
41	5	0,40	1,00	47	5	0,90	2,00

Tabela B.1 Cont. .

Item	alternativa	Disc. (a)	Dif. (b)	Item	alternativa	Disc. (a)	Dif. (b)
42	1	-0,50	-2,00	48	1	-1,10	-1,00
42	2	-0,20	-1,00	48	2	-0,80	0,00
42	3	0,10	0,00	48	3	0,40	1,00
42	4	0,20	1,00	48	4	0,60	2,00
42	5	0,40	2,00	48	5	0,90	3,00
43	1	-1,10	-1,00	49	1	-1,10	-2,50
43	2	-0,80	0,00	49	2	-0,80	-1,50
43	3	0,40	1,00	49	3	0,40	-0,50
43	4	0,60	2,00	49	4	0,60	1,20
43	5	0,90	3,00	49	5	0,90	2,50
44	1	-1,10	-2,50	50	1	-1,10	-1,50
44	2	-0,80	-1,50	50	2	-0,80	-0,50
44	3	0,40	-0,50	50	3	0,40	1,50
44	4	0,60	1,20	50	4	0,60	2,50
44	5	0,90	2,50	50	5	0,90	3,50



Tabela B.2 *Resultados da Simulação dos parâmetros dos itens (irrestritos) do Modelo de Resposta Nominal*

Item	Parâmetro									
	dificuldade ( $\delta$ )					discriminação ( $\alpha$ )				
	Valor	MVM		BM		Valor	MVM		BM	
		Est.	EP	Est.	EP		Est.	EP	Est.	EP
1	-1,250	-1,293	0,093	-1,288	0,092	-0,250	-0,232	0,081	-0,229	0,081
1	-2,050	-2,024	0,087	-2,019	0,087	-0,550	-0,507	0,078	-0,503	0,077
1	-2,250	-2,291	0,086	-2,285	0,086	-1,250	-1,187	0,080	-1,182	0,079
1	-1,300	-1,271	0,093	-1,265	0,093	-1,700	-1,706	0,088	-1,699	0,087
2	-1,000	-0,898	0,062	-0,896	0,062	-0,250	-0,307	0,058	-0,305	0,058
2	-1,500	-1,462	0,058	-1,460	0,058	-0,550	-0,600	0,056	-0,598	0,056
2	-1,000	-0,952	0,062	-0,949	0,062	-1,250	-1,268	0,063	-1,265	0,063
2	0,400	0,343	0,083	0,346	0,083	-1,700	-1,618	0,080	-1,614	0,079
3	-0,750	-0,703	0,047	-0,702	0,047	-0,250	-0,310	0,047	-0,310	0,046
3	-0,950	-0,937	0,045	-0,936	0,045	-0,550	-0,538	0,046	-0,536	0,046
3	0,250	0,272	0,060	0,274	0,060	-1,250	-1,226	0,061	-1,224	0,061
3	2,100	2,174	0,128	2,173	0,128	-1,700	-1,626	0,105	-1,620	0,105
4	-1,125	-1,267	0,072	-1,265	0,072	-0,250	-0,278	0,067	-0,276	0,067
4	-1,775	-1,913	0,068	-1,911	0,068	-0,550	-0,543	0,064	-0,541	0,064
4	-1,275	-1,406	0,071	-1,402	0,071	-1,250	-1,333	0,071	-1,329	0,070
4	0,500	0,337	0,101	0,340	0,101	-1,700	-1,736	0,093	-1,731	0,093
5	-0,875	-0,861	0,053	-0,861	0,053	-0,250	-0,249	0,052	-0,248	0,051
5	-1,425	-1,431	0,050	-1,430	0,050	-0,550	-0,584	0,049	-0,583	0,049
5	0,125	0,049	0,065	0,051	0,065	-1,250	-1,251	0,065	-1,248	0,065
5	2,200	2,244	0,153	2,242	0,153	-1,700	-1,808	0,118	-1,801	0,117
6	-1,250	-1,252	0,095	-1,248	0,094	-0,250	-0,195	0,084	-0,191	0,083
6	-2,050	-2,141	0,088	-2,136	0,088	-0,550	-0,491	0,079	-0,487	0,079
6	-2,250	-2,284	0,088	-2,278	0,087	-1,250	-1,181	0,081	-1,175	0,081
6	-1,300	-1,392	0,094	-1,386	0,094	-1,700	-1,615	0,088	-1,609	0,088
7	-1,000	-0,913	0,061	-0,911	0,061	-0,250	-0,223	0,058	-0,222	0,058
7	-1,500	-1,369	0,057	-1,367	0,057	-0,550	-0,567	0,056	-0,565	0,056
7	-1,000	-0,935	0,061	-0,932	0,061	-1,250	-1,244	0,063	-1,241	0,063
7	0,400	0,520	0,087	0,523	0,087	-1,700	-1,766	0,082	-1,762	0,082
8	-0,700	-0,692	0,046	-0,691	0,046	-0,200	-0,116	0,046	-0,116	0,046
8	-0,500	-0,505	0,048	-0,504	0,048	-0,900	-0,897	0,051	-0,895	0,051
8	-0,100	-0,105	0,052	-0,103	0,052	-1,000	-0,999	0,055	-0,997	0,055
8	1,400	1,295	0,084	1,296	0,084	-1,400	-1,315	0,081	-1,312	0,081
9	-1,000	-0,921	0,069	-0,920	0,069	-0,200	-0,109	0,063	-0,108	0,063
9	-1,850	-1,767	0,063	-1,765	0,063	-0,900	-0,791	0,060	-0,789	0,060
9	-1,390	-1,301	0,065	-1,298	0,065	-1,000	-0,896	0,063	-0,893	0,063
9	0,000	-0,008	0,082	-0,005	0,082	-1,400	-1,265	0,079	-1,261	0,079
10	-0,800	-0,729	0,049	-0,728	0,049	-0,200	-0,209	0,047	-0,208	0,047
10	-0,750	-0,703	0,049	-0,701	0,049	-0,900	-0,911	0,051	-0,909	0,051
10	-0,300	-0,296	0,053	-0,294	0,053	-1,000	-1,006	0,055	-1,003	0,055
10	1,400	1,398	0,093	1,399	0,093	-1,400	-1,413	0,086	-1,409	0,086
11	-1,100	-0,983	0,089	-0,980	0,089	-0,200	-0,073	0,078	-0,070	0,077
11	-2,300	-2,161	0,080	-2,157	0,080	-0,900	-0,739	0,073	-0,735	0,072
11	-2,100	-1,999	0,080	-1,995	0,080	-1,000	-0,856	0,074	-0,853	0,074
11	-1,400	-1,320	0,085	-1,316	0,085	-1,400	-1,225	0,079	-1,221	0,079

Tabela B.2 Cont. .

Item	Parâmetro									
	dificuldade ( $\delta$ )					discriminação ( $\alpha$ )				
	Valor	MVM		BM		Valor	MVM		BM	
		Est.	EP	Est.	EP		Est.	EP	Est.	EP
12	-0,900	-0,925	0,064	-0,923	0,064	-0,200	-0,257	0,059	-0,255	0,058
12	-1,400	-1,454	0,060	-1,451	0,060	-0,900	-0,866	0,058	-0,864	0,058
12	-1,100	-1,134	0,062	-1,131	0,062	-1,000	-1,028	0,061	-1,025	0,061
12	0,000	-0,072	0,076	-0,069	0,076	-1,400	-1,458	0,074	-1,454	0,074
13	-0,700	-0,735	0,047	-0,735	0,047	-0,200	-0,278	0,046	-0,277	0,046
13	-0,500	-0,577	0,048	-0,575	0,048	-0,900	-0,955	0,051	-0,953	0,051
13	-0,100	-0,145	0,053	-0,144	0,053	-1,000	-0,965	0,055	-0,963	0,055
13	1,400	1,318	0,087	1,319	0,087	-1,400	-1,444	0,081	-1,441	0,081
14	-1,000	-0,973	0,072	-0,971	0,072	-0,200	-0,190	0,064	-0,188	0,064
14	-1,850	-1,830	0,066	-1,828	0,066	-0,900	-0,909	0,062	-0,906	0,062
14	-1,390	-1,404	0,068	-1,401	0,068	-1,000	-1,069	0,065	-1,066	0,065
14	0,000	-0,062	0,085	-0,059	0,085	-1,400	-1,385	0,081	-1,381	0,081
15	-1,250	-1,274	0,061	-1,272	0,061	-0,500	-0,530	0,056	-0,528	0,056
15	-1,800	-1,813	0,058	-1,811	0,058	-0,800	-0,824	0,055	-0,822	0,055
15	0,000	0,039	0,078	0,042	0,078	-1,600	-1,630	0,074	-1,626	0,074
15	2,350	2,256	0,180	2,253	0,180	-2,100	-1,885	0,143	-1,876	0,142
16	-2,000	-2,044	0,132	-2,029	0,131	-0,500	-0,515	0,101	-0,504	0,101
16	-2,800	-2,891	0,128	-2,876	0,127	-0,800	-0,876	0,099	-0,864	0,098
16	-3,000	-3,088	0,127	-3,072	0,126	-1,600	-1,702	0,103	-1,689	0,102
16	-1,900	-2,013	0,133	-1,995	0,132	-2,100	-2,196	0,110	-2,182	0,109
17	-1,500	-1,485	0,076	-1,482	0,076	-0,500	-0,489	0,067	-0,486	0,067
17	-2,000	-2,001	0,074	-1,997	0,073	-0,800	-0,836	0,066	-0,833	0,066
17	-1,400	-1,399	0,077	-1,395	0,077	-1,600	-1,673	0,074	-1,667	0,074
17	0,200	0,077	0,101	0,082	0,101	-2,100	-1,970	0,093	-1,964	0,092
18	-1,000	-0,946	0,051	-0,946	0,051	-0,500	-0,383	0,050	-0,382	0,050
18	-1,200	-1,199	0,050	-1,198	0,050	-0,800	-0,739	0,050	-0,737	0,050
18	0,200	0,262	0,068	0,264	0,068	-1,600	-1,581	0,068	-1,577	0,068
18	2,300	2,070	0,137	2,069	0,136	-2,100	-1,882	0,112	-1,876	0,111
19	-1,750	-1,779	0,091	-1,774	0,091	-0,500	-0,501	0,076	-0,497	0,076
19	-2,400	-2,503	0,088	-2,497	0,088	-0,800	-0,909	0,075	-0,904	0,075
19	-1,780	-1,855	0,091	-1,848	0,091	-1,600	-1,684	0,082	-1,678	0,082
19	0,250	0,035	0,122	0,042	0,122	-2,100	-2,170	0,106	-2,162	0,105
20	-1,250	-1,255	0,061	-1,254	0,060	-0,500	-0,509	0,056	-0,507	0,056
20	-1,800	-1,799	0,058	-1,797	0,058	-0,800	-0,773	0,055	-0,770	0,055
20	0,000	-0,024	0,077	-0,022	0,077	-1,600	-1,624	0,074	-1,619	0,074
20	2,350	2,349	0,185	2,345	0,184	-2,100	-2,250	0,134	-2,240	0,133
21	-2,000	-1,974	0,123	-1,963	0,122	-0,500	-0,390	0,098	-0,381	0,098
21	-2,800	-2,730	0,119	-2,718	0,118	-0,800	-0,770	0,096	-0,761	0,096
21	-3,000	-2,935	0,118	-2,922	0,117	-1,600	-1,520	0,099	-1,509	0,099
21	-1,900	-1,816	0,124	-1,802	0,124	-2,100	-2,001	0,107	-1,989	0,106
22	-0,600	-0,631	0,050	-0,631	0,050	-0,200	-0,116	0,050	-0,115	0,050
22	-0,800	-0,829	0,049	-0,829	0,048	-0,300	-0,149	0,049	-0,149	0,048
22	-0,600	-0,578	0,051	-0,577	0,051	-0,600	-0,447	0,051	-0,446	0,051
22	0,200	0,202	0,062	0,203	0,062	-0,900	-0,817	0,061	-0,816	0,061

Tabela B.2 Cont. .

Item	Parâmetro									
	dificuldade ( $\delta$ )					discriminação ( $\alpha$ )				
	Valor	MVM		BM		Valor	MVM		BM	
		Est.	EP	Est.	EP		Est.	EP	Est.	EP
23	-0,400	-0,397	0,044	-0,396	0,044	-0,200	-0,181	0,044	-0,181	0,044
23	-0,500	-0,499	0,043	-0,498	0,043	-0,300	-0,281	0,044	-0,280	0,043
23	0,000	-0,007	0,048	-0,006	0,048	-0,600	-0,600	0,049	-0,599	0,049
23	1,100	1,038	0,069	1,039	0,069	-0,900	-0,900	0,066	-0,898	0,066
24	-0,700	-0,692	0,053	-0,692	0,053	-0,200	-0,193	0,052	-0,192	0,052
24	-0,950	-0,954	0,051	-0,953	0,051	-0,300	-0,343	0,050	-0,342	0,050
24	-0,760	-0,792	0,052	-0,791	0,052	-0,600	-0,587	0,052	-0,585	0,052
24	0,250	0,164	0,065	0,166	0,065	-0,900	-0,935	0,064	-0,933	0,064
25	-0,500	-0,458	0,046	-0,458	0,046	-0,200	-0,146	0,045	-0,146	0,045
25	-0,750	-0,722	0,043	-0,722	0,043	-0,300	-0,269	0,044	-0,269	0,044
25	-0,100	-0,089	0,050	-0,089	0,050	-0,600	-0,569	0,050	-0,568	0,050
25	1,150	1,110	0,074	1,110	0,074	-0,900	-0,844	0,071	-0,842	0,071
26	-0,800	-0,822	0,063	-0,820	0,063	-0,200	-0,230	0,060	-0,229	0,060
26	-1,100	-1,200	0,060	-1,198	0,060	-0,300	-0,431	0,058	-0,429	0,057
26	-1,200	-1,241	0,060	-1,239	0,060	-0,600	-0,702	0,058	-0,700	0,058
26	-0,700	-0,754	0,064	-0,751	0,064	-0,900	-1,004	0,063	-1,002	0,063
27	-0,600	-0,561	0,050	-0,560	0,050	-0,200	-0,161	0,049	-0,161	0,049
27	-0,800	-0,756	0,048	-0,756	0,048	-0,300	-0,245	0,048	-0,245	0,048
27	-0,600	-0,539	0,050	-0,538	0,050	-0,600	-0,556	0,051	-0,555	0,051
27	0,200	0,224	0,061	0,225	0,061	-0,900	-0,844	0,060	-0,842	0,060
28	-0,400	-0,420	0,044	-0,420	0,044	-0,200	-0,194	0,044	-0,194	0,044
28	-0,500	-0,539	0,043	-0,538	0,043	-0,300	-0,279	0,044	-0,278	0,043
28	0,000	-0,013	0,049	-0,012	0,049	-0,600	-0,639	0,050	-0,637	0,050
28	1,100	1,056	0,070	1,057	0,070	-0,900	-0,855	0,067	-0,853	0,067
29	-1,600	-1,535	0,083	-1,531	0,083	-0,400	-0,400	0,074	-0,396	0,073
29	-2,400	-2,357	0,080	-2,353	0,079	-0,800	-0,726	0,071	-0,722	0,071
29	-1,660	-1,622	0,083	-1,617	0,083	-1,700	-1,571	0,079	-1,565	0,078
29	0,250	0,333	0,120	0,338	0,120	-2,100	-2,063	0,105	-2,056	0,104
30	-1,200	-1,195	0,060	-1,193	0,059	-0,400	-0,357	0,056	-0,356	0,056
30	-1,800	-1,779	0,057	-1,777	0,056	-0,800	-0,778	0,055	-0,775	0,055
30	0,250	0,240	0,082	0,243	0,082	-1,700	-1,751	0,078	-1,747	0,077
30	2,350	2,453	0,190	2,448	0,189	-2,100	-2,221	0,139	-2,210	0,138
31	-1,800	-1,471	0,119	-1,462	0,118	-0,400	-0,158	0,098	-0,152	0,098
31	-2,800	-2,595	0,111	-2,585	0,110	-0,800	-0,652	0,094	-0,644	0,093
31	-3,000	-2,820	0,111	-2,809	0,110	-1,700	-1,568	0,097	-1,558	0,097
31	-1,900	-1,766	0,116	-1,754	0,116	-2,100	-1,905	0,104	-1,895	0,104
32	-1,400	-1,455	0,078	-1,452	0,078	-0,400	-0,388	0,069	-0,385	0,069
32	-2,000	-2,095	0,075	-2,091	0,075	-0,800	-0,851	0,068	-0,847	0,068
32	-1,300	-1,389	0,079	-1,383	0,079	-1,700	-1,737	0,077	-1,732	0,077
32	0,200	0,056	0,103	0,061	0,103	-2,100	-2,103	0,095	-2,097	0,095
33	-1,000	-1,069	0,053	-1,068	0,053	-0,400	-0,509	0,051	-0,508	0,051
33	-1,200	-1,296	0,051	-1,294	0,051	-0,800	-0,858	0,051	-0,856	0,051
33	0,400	0,496	0,077	0,499	0,077	-1,700	-1,839	0,075	-1,834	0,074
33	2,300	2,360	0,159	2,357	0,158	-2,100	-2,338	0,120	-2,329	0,119

Tabela B.2 Cont. .

Item	Parâmetro									
	dificuldade ( $\delta$ )					discriminação ( $\alpha$ )				
	Valor	MVM		BM		Valor	MVM		BM	
		Est.	EP	Est.	EP		Est.	EP	Est.	EP
34	-1,600	-1,724	0,089	-1,719	0,089	-0,400	-0,399	0,078	-0,395	0,077
34	-2,400	-2,513	0,086	-2,508	0,085	-0,800	-0,727	0,075	-0,722	0,075
34	-1,660	-1,813	0,089	-1,807	0,089	-1,700	-1,579	0,083	-1,573	0,082
34	0,250	0,152	0,123	0,158	0,123	-2,100	-1,965	0,108	-1,958	0,108
35	-1,200	-1,261	0,062	-1,259	0,062	-0,400	-0,511	0,057	-0,509	0,057
35	-1,800	-1,879	0,059	-1,877	0,059	-0,800	-0,882	0,056	-0,880	0,056
35	0,250	0,144	0,082	0,147	0,082	-1,700	-1,751	0,078	-1,746	0,077
35	2,350	2,479	0,199	2,474	0,198	-2,100	-2,296	0,143	-2,285	0,142
36	-1,100	-1,025	0,070	-1,023	0,070	-0,300	-0,245	0,065	-0,244	0,065
36	-1,600	-1,578	0,066	-1,575	0,066	-0,600	-0,522	0,062	-0,520	0,062
36	-1,500	-1,466	0,066	-1,464	0,066	-0,700	-0,573	0,063	-0,570	0,063
36	-1,100	-1,128	0,069	-1,125	0,069	-0,900	-0,875	0,066	-0,872	0,066
37	-0,800	-0,829	0,056	-0,828	0,056	-0,300	-0,295	0,053	-0,294	0,053
37	-1,000	-1,040	0,054	-1,038	0,054	-0,600	-0,610	0,053	-0,608	0,053
37	-0,800	-0,848	0,056	-0,847	0,056	-0,700	-0,747	0,055	-0,745	0,055
37	-0,200	-0,219	0,063	-0,217	0,063	-0,900	-0,924	0,062	-0,922	0,062
38	-0,500	-0,487	0,045	-0,486	0,045	-0,300	-0,289	0,044	-0,289	0,044
38	-0,400	-0,381	0,046	-0,380	0,046	-0,600	-0,616	0,047	-0,614	0,047
38	-0,100	-0,146	0,048	-0,145	0,048	-0,700	-0,722	0,049	-0,721	0,049
38	0,700	0,783	0,064	0,784	0,064	-0,900	-0,903	0,064	-0,901	0,063
39	-0,950	-0,903	0,059	-0,901	0,059	-0,300	-0,335	0,055	-0,334	0,055
39	-1,300	-1,259	0,056	-1,257	0,056	-0,600	-0,606	0,054	-0,604	0,054
39	-1,010	-0,916	0,059	-0,914	0,058	-0,700	-0,755	0,057	-0,753	0,057
39	-0,250	-0,319	0,065	-0,317	0,065	-0,900	-0,937	0,064	-0,934	0,063
40	-0,650	-0,539	0,045	-0,539	0,045	-0,300	-0,264	0,045	-0,263	0,044
40	-0,600	-0,518	0,045	-0,517	0,045	-0,600	-0,563	0,046	-0,562	0,046
40	-0,250	-0,050	0,050	-0,049	0,050	-0,700	-0,626	0,051	-0,624	0,051
40	0,650	0,777	0,065	0,779	0,065	-0,900	-0,966	0,064	-0,964	0,064
41	-1,100	-1,178	0,072	-1,176	0,072	-0,300	-0,265	0,066	-0,263	0,066
41	-1,600	-1,627	0,069	-1,624	0,069	-0,600	-0,594	0,064	-0,591	0,064
41	-1,500	-1,600	0,069	-1,597	0,069	-0,700	-0,655	0,065	-0,653	0,064
41	-1,100	-1,164	0,072	-1,161	0,072	-0,900	-0,837	0,068	-0,834	0,068
42	-0,800	-0,788	0,056	-0,787	0,056	-0,300	-0,302	0,054	-0,301	0,054
42	-1,000	-1,064	0,054	-1,063	0,054	-0,600	-0,529	0,052	-0,528	0,052
42	-0,800	-0,863	0,055	-0,861	0,055	-0,700	-0,619	0,054	-0,618	0,054
42	-0,200	-0,315	0,061	-0,313	0,061	-0,900	-0,840	0,060	-0,838	0,060
43	-1,100	-1,116	0,051	-1,115	0,051	-0,300	-0,206	0,050	-0,205	0,050
43	-0,700	-0,753	0,054	-0,750	0,054	-1,500	-1,446	0,062	-1,443	0,062
43	0,100	0,148	0,067	0,151	0,067	-1,700	-1,726	0,073	-1,722	0,072
43	1,600	1,495	0,108	1,496	0,107	-2,000	-2,029	0,100	-2,023	0,100
44	-1,550	-1,600	0,109	-1,591	0,108	-0,300	-0,411	0,083	-0,404	0,083
44	-2,950	-3,011	0,103	-3,001	0,102	-1,500	-1,642	0,085	-1,633	0,085
44	-2,030	-2,068	0,106	-2,058	0,105	-1,700	-1,803	0,090	-1,794	0,090
44	-0,500	-0,598	0,123	-0,588	0,123	-2,000	-2,099	0,108	-2,088	0,108

Tabela B.2 Cont. .

Item	Parâmetro									
	dificuldade ( $\delta$ )					discriminação ( $\alpha$ )				
	Valor	MVM		BM		Valor	MVM		BM	
		Est.	EP	Est.	EP		Est.	EP	Est.	EP
45	-1,250	-1,148	0,055	-1,147	0,055	-0,300	-0,195	0,053	-0,194	0,053
45	-1,050	-0,986	0,056	-0,983	0,056	-1,500	-1,480	0,063	-1,476	0,062
45	-0,150	-0,123	0,066	-0,120	0,066	-1,700	-1,672	0,072	-1,668	0,072
45	1,500	1,542	0,116	1,544	0,116	-2,000	-1,932	0,109	-1,925	0,108
46	-1,700	-1,485	0,152	-1,467	0,151	-0,300	-0,234	0,108	-0,221	0,107
46	-3,700	-3,537	0,141	-3,516	0,139	-1,500	-1,526	0,106	-1,510	0,105
46	-3,300	-3,143	0,142	-3,122	0,140	-1,700	-1,701	0,108	-1,685	0,107
46	-2,400	-2,288	0,145	-2,267	0,144	-2,000	-2,032	0,113	-2,015	0,112
47	-1,400	-1,517	0,088	-1,512	0,088	-0,300	-0,359	0,072	-0,355	0,072
47	-2,200	-2,286	0,084	-2,280	0,084	-1,500	-1,512	0,077	-1,506	0,076
47	-1,600	-1,705	0,087	-1,698	0,087	-1,700	-1,758	0,081	-1,751	0,081
47	-0,400	-0,498	0,101	-0,491	0,100	-2,000	-1,995	0,095	-1,987	0,094
48	-1,100	-1,093	0,051	-1,092	0,051	-0,300	-0,264	0,050	-0,263	0,050
48	-0,700	-0,709	0,055	-0,706	0,054	-1,500	-1,492	0,062	-1,489	0,061
48	0,100	0,042	0,064	0,045	0,064	-1,700	-1,704	0,070	-1,700	0,070
48	1,600	1,504	0,107	1,506	0,107	-2,000	-1,948	0,101	-1,943	0,101
49	-1,550	-1,651	0,115	-1,641	0,114	-0,300	-0,326	0,088	-0,320	0,087
49	-2,950	-3,138	0,108	-3,127	0,108	-1,500	-1,584	0,089	-1,574	0,089
49	-2,030	-2,259	0,111	-2,248	0,110	-1,700	-1,717	0,093	-1,708	0,093
49	-0,500	-0,636	0,129	-0,625	0,129	-2,000	-1,983	0,113	-1,973	0,113
50	-1,250	-1,178	0,054	-1,178	0,054	-0,300	-0,245	0,052	-0,244	0,052
50	-1,050	-0,996	0,056	-0,994	0,055	-1,500	-1,376	0,061	-1,373	0,061
50	-0,150	-0,122	0,066	-0,119	0,066	-1,700	-1,609	0,070	-1,605	0,070
50	1,500	1,528	0,115	1,530	0,115	-2,000	-1,986	0,104	-1,980	0,104

---

## Referências Bibliográficas

---

- [1] Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- [2] Albert, J. H. (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*, 17, 251 - 269.
- [3] Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. New York: North-Holand Publishing Company.
- [4] Andersen, E. B. (1997). The Rating Scale Model. In *Handbook of Modern Item Response Theory*. W. J. van der Linden and R. K. Hambleton Eds. New York: Springer - Verlag.
- [5] Andrade, D. F. (2001). Desempenhos de Grupos de Alunos por Intermédio da Teoria de Resposta ao Item. *Estudos em Avaliação Educacional*, 23, 31-70, Fundação Carlos Chagas.
- [6] Andrade, D. F. , Tavares, H. R. & Valle, R. C. (2000). Introdução à Teoria da resposta ao Item: Conceitos e Aplicações. 14<sup>o</sup> SINAPE: Caxambú, MG.
- [7] Azevedo, C. L. N. (2000 a). Métodos Estatísticos para Avaliação Educacional: Teoria Clássica de Medidas e Teoria da Resposta ao Item. *UFC e Programa Institucional de Bolsas de Iniciação Científica - PIBIC/CNPq - Relatório Final*.
- [8] Azevedo, C. L. N. (2000 b). Métodos Estatísticos para Avaliação Educacional: Teoria da Resposta ao Item e Planejamento para Montagem de provas. *UFC e Programa Institucional de Bolsas de Iniciação Científica - PIBIC/CNPq - Relatório Final*.
- [9] Azevedo, C. L. N. (2002). Métodos de Estimação na Teoria da Resposta ao Item: Máxima Verossimilhança versus Métodos Bayesianos. Trabalho apresentado ao 15<sup>o</sup> SINAPE: Águas de Lindóia, SP.
- [10] Baker, F. B. (1990). Some observations on the metric of PC-Bilog results. *Applied Psychological Measurement*, 14, 19 - 150.
- [11] Baker, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.

- [12] Bernardo, J. M. & Smith, A. F. M. (2001). *Bayesian Theory*. John Wiley & Sons.
- [13] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F. M. Lord & M. R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- [14] Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- [15] Bock, R. D. (1997). The Nominal Categories Model. In *Handbook of Modern Item Response Theory*. W. J. van der Linden and R. K. Hambleton Eds. New York: Springer - Verlag.
- [16] Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm. *Psychometrika*, 46, 433-459.
- [17] Bock, R. D. & Lieberman, M. (1970). Fitting a response model for  $n$  dichotomously scored items. *Psychometrika*, 35, 179-197.
- [18] Bock, R. D. & Zimowski, M. F. (1997). Multiple Group IRT. In *Handbook of Modern Item Response Theory*. W. J. van der Linden and R. K. Hambleton Eds. New York: Springer - Verlag.
- [19] Bolt, D. M. & Lall, V. F. (2001). Estimation of Compensatory and Noncompensatory Multidimensional Item Response Models Using Markov Chain Monte Carlo. *Submetido ao Division D1: Educational Measurement, Psychometrics, and Assessment*.
- [20] Boomsma, A., van Duijn, M. A. J. & Snijders, T. A. B. (2001). *Essays on Item Response Theory*. Springer - Verlag New York, Inc.
- [21] Chang, Hua-Hua (1996). The Asymptotic Normality of the Latent Trait for Polytomous IRT Models. *Psychometrika*, 61, 445-463.
- [22] Chang, Hua-Hua & Stout, W. (1993). The Asymptotic Normality of the Latent Trait in an IRT Model. *Psychometrika*, 58, 7 - 52.
- [23] Cordeiro, G. (1992). Introdução à Teoria da Verossimilhança. 10º SINAPE.
- [24] Cornfield, J. (1969). The Bayesian outlook and its applications. *Biometrics*, 25, 617 - 657.
- [25] Davey, T., Nering, M. L. & Thompson, Tony (1997). Realistic Simulation of Item Response Data. *ACT Research Report Series* (www.act.org).
- [26] de Finetti, B. (1974). Bayesianism: Its unifying role for both the foundations and applications of Statistics. *International Statistical Review*, 42, 117 - 130.

- [27] Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- [28] Doornik, J. A. (2000). Objected-Oriented Matrix Programming using Ox 2.0. London: Timberlake Consultants Ltd and Oxford, [www.nuff.ox.ac.uk/Users/Doornik](http://www.nuff.ox.ac.uk/Users/Doornik).
- [29] Edwards, W., Lindman, H. & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.
- [30] Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Londres: Chapman & Hall.
- [31] Garland, R. (1991). The Mid-Point on a Rating Scale: Is it Desirable? *Marketing Bulletin*, 2, 66-70.
- [32] Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Transaction on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- [33] Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman & Hall.
- [34] Ghosh, M., Ghosh, A., Chen, Ming-Hui & Agresti, A. (1998). Bayesian Estimation fo Item Response Models. Comunicação Pessoal.
- [35] Harville, D. A. (1997). *Matrix Algebra from a statistician's perspective*. New York: Springer.
- [36] Harwell, M. R. & Janosky, J. E. (1991). An empirical study of the effects of sparse data sets and varying prior variances on item parameter estimation in Bilog. *Applied Psychological Measurement*, 15, 279-291.
- [37] Harwell, M. R., Baker, F. B. & Zwarts, M. (1988). Item Parameter Estimation via Marginal Maximum Likelihood and an EM algorithm: A didatic. *Journal of Educational Statistics*, 13, 243-271.
- [38] Hastings, W. K. (1970). Monte Carlo simulation methods using Markov Chains and their applications. *Biometrika*, 57, 97-109.
- [39] Hildebrand, F. B. (1956). *Introduction to Numerical Analysis*. New-York: McGraw-Hill.
- [40] Hsu, Y., Ackerman, Terry & Fan, M. (1999). The Relationship Between the Bock-Aitkin Procedure and the EM Algorithm for IRT Model Estimation. *ACT Research Report Series* ([www.act.org](http://www.act.org)).



- [41] Houtink, H. & Boomsma, A. (1996). Statistical Inference Based on Latent Ability Estimates. *Psychometrika*, 61, 313-330.
- [42] Issac, E. & Keller, H.B. (1966). *Analysis of Numerical Methods*. New York: Wiley & Sons.
- [43] Jones, H. D. & Nediak, M. (2000). Item Parameter calibration of LSAT Items Using MCMC Approximation of Bayes Posterior Distributions. *Rutcor Research Report, RRR 7-2000, Fevereiro, 2000*. Rutgers Center for Operational Research, Rutgers University (<http://rutcor.rutgers.edu/rrr>).
- [44] Jorgensen, B. & Labouriau, R. S. (1992). *Famílias exponenciais e inferência teórica*. Rio de Janeiro: Conselho Nacional de Desenvolvimento Científico e Tecnológico, Instituto de Matemática Pura e Aplicada.
- [45] Kim, J. K. & Nicewander, W. A. (1993). Ability Estimation for Conventional Tests. *Psychometrika*, 58, 587-599.
- [46] Kolakowski, D. & Bock, R. D. (1973 a). *NORMOG: Maximum Likelihood item analysis and test scoring : normal ogive model*. Chicago : International Educational services.
- [47] Kolakowski, D. & Bock, R. D. (1973 b). *LOGOG: Maximum Likelihood item analysis and test scoring: logistic model for multiple item responses*. Chicago: International Educational services.
- [48] Kolen, M. J. & Brennan, R. L. (1995). *Test Equating - Methods and Practices*. New York.
- [49] Lindley, D. V. (1970 a). *Introduction to Probability and Statistics: From a Bayesian Viewpoint. Part 1 : Probability*. Cambridge, England: Cambridge Univeristy Press.
- [50] Lindley, D. V. (1970 b). *Introduction to Probability and Statistics: From a Bayesian Viewpoint. Part 2: Inference*. Cambridge, England: Cambridge Univeristy Press.
- [51] Lindley, D. V. & Smith, A. F. M. (1972). Bayesian estimates for the linear model. *Journal of the Royal Statistical Society*, 34, 1-41.
- [52] Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.
- [53] Lord, F. M. (1975). *Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters*, (Research Bulletin RB-75-3). Princeton, NJ: Educational Testing Service.
- [54] Lord, F. M. (1983). Unbiased estimation of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, 48, 233-445.

- [55] Lord, F. M. (1984). *Maximum Likelihood and Bayesian Parameter Estimation in Item Response Theory* (Research Report. No. RR-84-30-ONR). Princeton, NJ: Educational Testing Service.
- [56] Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.
- [57] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- [58] Matos, G. S. (2001). *Teoria da Resposta ao Item: Uma Proposta de Modelo Multivariado*. Dissertação de Mestrado. Universidade Federal de Pernambuco - UFPE - Departamento de Estatística.
- [59] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state space calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1091.
- [60] Mislevy, J. M. (1984). Estimating Latent Distributions. *Psychometrika*, Vol. 49, 359 - 381.
- [61] Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177-195.
- [62] Mislevy, R. J. (1992). *Linking Educational Assessments : concepts, issues, methods and prospects*. Princeton Educational Testing Service.
- [63] Mislevy, R. J. & Bock, R. D. (1982). Biweight Estimates of Latent Ability. *Educational and Psychological Measurement*, 42, 725-737.
- [64] Mislevy, R. J. & Bock, R. D. (1990). *BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models*. Chicago: Scientific Software, Inc.
- [65] Mislevy, R. J. & Stocking, M. L. (1989). A Consumer's Guide to LOGISTIC and BILOG. *Applied Psychological Measurement*, 13, 57-75.
- [66] Mosteller, F. & Tukey, J (1977). *Exploratory Data Analysis and Regression*. Reading, Mass.: Addison-Wesley.
- [67] Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- [68] Muraki, E. & Bock, R. D.(1997). *PARSCALE: IRT Based Test Scoring and Item Analysis fo Graded Open-Ended Exercises and Performance Tasks*. Chicago: Scientific Siftware, Inc.

- [69] Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, 16 (1), 1-32.
- [70] Nojosa, R. T. (2001). *Modelos Multidimensionais na Teoria da Resposta ao Item*. Dissertação de Mestrado. Universidade Federal de Pernambuco - UFPE - Departamento de Estatística.
- [71] Novick, M. R. & Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.
- [72] O'Hagan, A. (1976). On posterior joint and marginal modes. *Biometrika*, 63, 329-333.
- [73] Patz, J. R. & Junker, B. W. (1999 a). A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models. *Journal of Educational and Behavioral Statistics*, Vol. 24, 146 - 178.
- [74] Patz, J. R. & Junker, B. W. (1999 b). Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing data, and Rated Responses. *Journal of Educational and Behavioral Statistics*, Vol. 24, 342-366.
- [75] Ramsay, J. O. (1975). Solving Implicit Equations in Psychometric Data Analysis. *Psychometrika*, 40, 337 - 360.
- [76] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley & Sons.
- [77] Sahu, S. K. (2001). Bayesian Estimation and Model Choice in Item Response Models. Comunicação Pessoal.
- [78] Samejima, F. A. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 17.
- [79] Samejima, F. A. (1997). Graded Response Model. In *Handbook of Modern Item Response Theory*. W. J. van der Linden e R. K. Hambleton Eds. New York : Springer - Verlag.
- [80] Sen, P.K. & Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction With Applications*. New York: Chapman & Hall.
- [81] Swaminathan, H. & Gifford, J. A. (1982). Bayesian Estimation in the Rasch model. *Journal of Educational Statistics*, 7, 175 - 192.
- [82] Swaminathan, H. & Gifford, J. A. (1985). Bayesian Estimation in the two-parameter logistic model. *Psychometrika*, 50, 349 - 364.

- [83] Swaminathan, H. & Gifford, J. A. (1986). Bayesian Estimation in the three-parameter logistic model. *Psychometrika*, 51, 589 - 601.
- [84] Stroud, A. H. and Secret, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, New Jersey: Prentice-Hall.
- [85] Tavares, H. R. *Teoria da Resposta ao Item para Dados Longitudinais*. Tese de Doutorado. São Paulo: IME/USP.
- [86] Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175-186.
- [87] Thissen, D. (1991). *Multilog User's Guide: Multiple, Categorical Item Analysis and Test Scoring Using Item Response Theory*. Chicago: Scientific Software, Inc.
- [88] Thissen, D. & Steinberg, L. (1984). A model for multiple choice items. *Psychometrika*, 49, 501 - 519.
- [89] Tierney, L. (1994). Exploring posterior distributions with Markov Chains. *Annals of Statistics*, 22, 1701-1762.
- [90] Tsutakawa, R. K. & Lin, H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika*, 51, 251-267.
- [91] van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- [92] Zimowski, M. F. , Muraki, E. , Mislevy, R. J. & Bock, R. D. (1996). *BILOG-MG: Multiple-Group IRT and Test Maintenance for Binary Items*. Chicago: Scientific Software, Inc.
- [93] Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54, 427-450.
- [94] Wollack, J. A, Bolt, D. M., Cohen, A. S. & Lee, Young-Sun (2000). Recovery of Item Parameters in the Nominal Response Model: A Comparison of Marginal Maximum Likelihood Estimation and Markov Chain Monte Carlo Estimation. *Trabalho apresentado no Encontro Nacional da Associação Americana de Pesquisa Educacional, Março, 2000*.
- [95] Wood, R. L., Wingersky & Lord, F. M. (1976). *LOGIST : A computer program for estimating examinee ability and item characteristic curve parameters*. Princeton (N.J.): Educational Testing Service, 1976.
- [96] Woodruff, D. & Hanson, B. A. (1996). Estimation for Item Response Models Using the EM Algorithm for Finite Mixtures, *ACT Research Report Series* ([www.act.org](http://www.act.org)).

- 
- [97] Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97 - 116.
- [98] Wright, B. D. & Panchapakesan, N. A. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 2-48.