Teoria da Resposta ao Item: Conceitos e Aplicações

Dalton Francisco de Andrade[1]

Heliton Ribeiro Tavares[2]

Raquel da Cunha Valle[3]

[1] Professor Titular do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará (UFC). *e-mail: dandrade@ufc.br*

[2] Professor do Departamento de Estatística da Universidade Federal do Pará (UFPA). *e-mail: heliton@ufpa.br*

[3] Estatístico da Fundação Carlos Chagas (FCC). *e-mail: rvalle@fcc.gov.br*

Para

Janete, Cristina e Fernando.

Regina e Henrique

Apresentação

A avaliação educacional passou a ser, embora tardiamente, um dos pontos privilegiados das políticas educacionais. Já são inúmeros os projetos de avaliação conduzidos por órgãos responsáveis pelos destinos da Educação em nosso país. Reclamava-se, porém, por uma metodologia mais sofisticada e precisa, que permitisse não só a avaliação pontual mas, sobretudo, a construção de escalas de habilidades que pudessem levar a um acompanhamento do progresso do conhecimento adquirido pelos alunos ao longo do tempo.

A Teoria Clássica, baseada em resultados obtidos em provas através de escores brutos ou padronizados, largamente utilizada até então, padece de várias limitações, como, por exemplo, ser dependente do conjunto de itens que compõem o instrumento de medida, limitando assim, a sua aplicabilidade.

A Teoria da Resposta ao Item (TRI), que vem sendo progressivamente introduzida em nosso meio, é um instrumento poderoso nos processos quantitativos de avaliação educacional, pelo fato de permitir, inclusive, a construção de escalas de habilidade calibradas. No entanto, a aplicabilidade da TRI tem encontrado algumas dificuldades, tanto do ponto de vista teórico, devido a problemas de difícil solução no campo da estimação, como do ponto de vista computacional.

O livro de Dalton F. Andrade, Heliton R. Tavares e Raquel C. Valle, vem ao encontro de uma real necessidade dos pesquisadores clarificando alguns pontos essenciais da teoria, trazendo um exemplo prático de aplicação em larga escala, como é o caso do Sistema de Avaliação do Rendimento Escolar do Estado de S.Paulo (SARESP).

Escrito de forma extremamente didática, não requerendo do leitor conhecimentos muito aprofundados do ponto de vista matemático-estatístico, com exceção de algumas partes dos capítulos de estimação, aborda os principais

modelos matemáticos utilizados, os problemas de estimação e equalização, e aponta os recursos computacionais adequados.

Certamente, o texto se tornará um referencial obrigatório para todos aqueles interessados em contribuir para o progresso dos aspectos quantitativos e metodológicos da Educação Brasileira.

Rubens Murillo Marques
Prof. Titular Estatística-Matemática da UNICAMP
Diretor Presidente da Fundação Carlos Chagas

Prefácio

A idéia de escrever um texto introdutório sobre a Teoria da Resposta ao Item – TRI, até agora tão pouco conhecida pelos especialistas em avaliação e pelos estatísticos no Brasil, surgiu da necessidade de se divulgar o potencial dessa teoria tanto no seu aspecto estatístico-matemático quanto na sua aplicação e interpretação na avaliação da aprendizagem e em outras áreas.

Nosso envolvimento com a TRI começou em 1996, com a análise dos dados gerados pela pesquisa AVEJU, da Secretaria de Estado da Educação de São Paulo, e continuou no Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo – SARESP e no Sistema de Avaliação da Educação Básica – SAEB do INEP/MEC. Esses dois sistemas de avaliação possuem a sua base metodológica fundamentada na TRI e são, atualmente, os grandes exemplos no Brasil da sua potencialidade.

Nossa maior preocupação foi a de escrever um texto que pudesse ser utilizado não só pelos estatísticos, mas também pelos especialistas em avaliação. O sucesso da TRI passa necessariamente pelo trabalho conjunto de especialistas dessas duas áreas. Devido a enorme abrangência da TRI. Nesse sentido, procuramos detalhar alguns pontos que achamos importantes.

Muito do material e idéias apresentadas nesse livro foram desenvolvidos durante o planejamento e a análise do SARESP e nos treinamentos que ministramos para técnicos da Secretaria de Estado da Educação de São Paulo, da Fundação para o Desenvolvimento da Educação - FDE e da Fundação Carlos Chagas, aos quais queremos agradecer a paciência e dedicação. Gostariamos também de expressar os nossos maiores agradecimentos a Yara Lúcia Espósito, Ruben Klein e Heraldo Vianna pelos longos papos e discussões sobre os aspectos teóricos e aplicados da TRI e a Profa. Rose Neubauer, Secretária de

Estado da Educação de São Paulo, pela utilização de parte dos resultados do SARESP.

Devido a enorme abrangência da TRI, procuramos detalhar os pontos que achamos mais interessantes para um texto introdutório e fornecer o maior número possível de referências bibliográficas que cobrissem os outros pontos.

Este trabalho foi parcialmente financiado pelo CNPq, pela CAPES, pelo Projeto Temático da FAPESP no. 96/01741-7 e pelo PRONEX no. 76.97.1081.00.

Fevereiro 2000

Dalton Francisco de Andrade
Heliton Ribeiro Tavares
Raquel da Cunha Valle

Conteúdo

Lista de Figuras

Capítulo **1**

Introdução

Resultados obtidos em provas, expressos apenas por seus escores brutos ou padronizados, têm sido tradicionalmente utilizados nos processos de avaliação e seleção de indivíduos. No entanto, os resultados encontrados dependem do particular conjunto de itens (questões) que compõem o instrumento de medida, ou seja, as análises e interpretações estão sempre associadas à prova como um todo, o que é a característica principal da Teoria Clássica das Medidas. Assim, torna-se inviável a comparação entre indivíduos que não foram submetidos às mesmas provas, ou pelo menos, ao que se denomina de formas paralelas de testes. Maiores detalhes sobre essa metodologia, incluindo sua fundamentação matemática, podem ser encontrados em Gulliksen (1950), Lord & Novick (1968) e Vianna (1987), entre outros.

Atualmente, em várias áreas do conhecimento, particularmente em avaliação educacional, vem crescendo o interesse na aplicação de técnicas derivadas da Teoria de Resposta ao Item – TRI, que propõe modelos para os traços latentes, ou seja, características do indivíduo que não podem ser observadas diretamente. Esse tipo de variável deve ser inferida a partir da observação de variáveis secundárias que estejam relacionadas a ela. O que esta metodologia sugere são formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item e seus traços latentes, proficiências ou habilidades na área de conhecimento avaliada.

Uma das grandes vantagens da TRI sobre a Teoria Clássica é que ela permite a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes. Isto porque uma das principais características da TRI é que ela tem como elementos centrais os itens, e não a prova como um todo.

Assim, várias questões de interesse prático na área da Educação podem

ser respondidas. É possível por exemplo, avaliar o desenvolvimento de uma determinada série de um ano para outro ou comparar o desempenho entre escolas públicas e privadas.

Os primeiros modelos de resposta ao item surgiram na década de 50, e eram modelos em que se considerava que uma única habilidade, de um único grupo, estava sendo medida por um teste onde os itens eram corrigidos de maneira dicotômica. Estes modelos foram primeiramente desenvolvidos na forma de uma função ogiva normal e, depois, foram descritos para uma forma matemática mais conveniente, e que vem sendo usada até então: a logística.

Lord (1952) foi o primeiro a desenvolver o modelo unidimensional de 2 parâmetros, baseado na distribuição normal acumulada (ogiva normal). Após algumas aplicações desse modelo, o próprio Lord sentiu a necessidade da incorporação de um parâmetro que tratasse do problema do acerto casual. Assim, surgiu o modelo de 3 parâmetros. Anos mais tarde, Birnbaum (1968) substituiu, em ambos os modelos, a função ogiva normal pela função logística, matematicamente mais conveniente, pois é uma função explícita dos parâmetros do item e de habilidade e não envolve integração. Independentemente do trabalho de Lord, Rasch (1960) propôs o modelo unidimensional de 1 parâmetro, expresso também como modelo de ogiva normal e, também mais tarde descrito por um modelo logístico por Wright (1968).

Samegima (1969) propôs o modelo de resposta gradual com o objetivo de obter mais informação das respostas dos indivíduos do que simplesmente se eles deram respostas corretas ou incorretas aos itens. Bock (1972), Andrich (1978), Masters (1982) e Muraki (1992) também propuseram modelos para mais de duas categorias de resposta, assumindo diferentes estruturas entre essas categorias.

Recentemente, Bock & Zimowski (1997) introduziram os modelos logísticos de 1, 2 e 3 parâmetros para duas ou mais populações de respondentes. A introdução desses modelos trouxe novas possibilidades para as comparações de rendimentos de duas ou mais populações submetidas a diferentes testes com itens comuns, conforme discutido em Hedges & Vevea (1997) e Andrade (1999), por exemplo.

Um ponto crítico na TRI é a estimação dos parâmetros envolvidos nos modelos, em particular quando necessita-se estimar tanto os parâmetros dos itens quanto as habilidades. Inicialmente, a estimação era feita através do

método da máxima verossimilhança conjunta que envolve um número muito grande de parâmetros a serem estimados simultaneamente e, consequentemente, grandes problemas computacionais. Em 1970, Bock & Lieberman introduziram o método da máxima verossimilhança marginal para a estimação dos parâmetros em duas etapas. Na primeira etapa estimam-se os parâmetros dos itens, assumindo-se uma certa distribuição para as habilidades. Na segunda etapa, assumindo os parâmetros dos itens conhecidos, estimam-se as habilidades. Apesar do avanço que esse método trouxe para o problema, ele requeria que todos os parâmetros dos itens fossem estimados simultaneamente. Em 1981, Bock & Aitkin propuseram uma modificação no método acima, utilizando o algoritmo EM de Dempster, Laird & Rubin (1977), de modo a permitir que os itens pudessem ter seus parâmetros estimados em separado, facilitando em muito o aspecto computacional do processo de estimação. Mais recentemente, métodos bayesianos foram propostos para, entre outras coisas, resolver o problema de estimação dos parâmetros dos itens respondidos corretamente ou incorretamente por todos os respondentes, e também o problema da estimação das habilidades dos respondentes que acertaram ou erraram todos os itens da prova.

Nas últimas décadas, a TRI vem tornando-se a técnica predominante no campo de testes em vários países. Aqui no Brasil, a TRI foi usada pela primeira vez em 1995 na análise dos dados do Sistema Nacional de Ensino Básico - SAEB. A introdução da TRI permitiu que os desempenhos de alunos de 4a. e 8a. séries do Ensino Fundamental e de 3a. série do Ensino Fundamental pudessem ser comparados e colocados em uma escala única de conhecimento. A partir dos resultados obtidos no SAEB, outras avaliações em larga escala, como por exemplo o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP, também foram planejadas e implemementadas de modo a serem analisadas através da TRI. Uma lista das principais aplicações da TRI no Brasil em avaliações educacionais pode ser encontrada em Andrade & Klein (1999).

O objetivo desse livro é introduzir os principais conceitos, modelos e resultados que podem ser obtidos a partir da aplicação da TRI. No Capítulo 2 são apresentados os modelos, com suas interpretações e suposições básicas. No Capítulo 3 discute-se o processo de estimação dos parâmetros dos itens e das habilidades dos respondentes pertencentes a uma única população. O

conceito de equalização e suas diferentes formas de obtenção são discutidos no Capítulo 4. Os métodos de estimação são retomados no Capítulo 5 com o modelo para duas ou mais populações. No Capítulo 6 discute-se a criação de escalas de habilidade e suas interpretações e uma aplicação a dados reais. No Capítulo 7 apresentam-se os principais recursos computacionais e no Capítulo 8 apresentam-se comentários sobre a utilização da TRI, inclusive em outras áreas, e possíveis tópicos para pesquisa. Por último, apresentam-se demonstrações de alguns dos resultados do Capítulo 3 no Apêndice e uma bibliografia com outras referências além daquelas citadas no texto, com o objetivo de fornecer ao leitor o maior número de informações sobre a TRI.

Os autores recomendam fortemente a leitura de Lord (1980) e Hambleton, Swaminathan & Rogers (1991) para maiores detalhes dos fundamentos e aplicações dessa teoria.

Capítulo **2**

Modelos Matemáticos

2.1 Introdução

A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item. Os vários modelos propostos na literatura dependem fundamentalmente de três fatores:

(i) da natureza do item — dicotômicos ou não dicotômicos;

(ii) do número de populações envolvidas — apenas uma ou mais de uma;

(iii) e da quantidade de traços latentes que está sendo medida — apenas um ou mais de um.

Nesse livro estaremos somente considerando modelos que avaliam apenas um traço latente ou habilidade, os chamados modelos unidimensionais. Modelos que consideram que mais de uma habilidade está sendo medida, os chamados modelos multidimensionais, podem ser encontrados em Linden & Hambleton (1997), por exemplo.

Na Seção 2.2 apresentaremos os modelos unidimensionais mais utilizados para um único grupo. Os modelos para dois ou mais grupos serão discutidos na Seção 2.3.

2.2 Modelos envolvendo um único grupo

Em primeiro lugar, é importante definir os conceitos de grupo e população, que serão largamente utilizados neste e nos demais capítulos. Quando usarmos o termo *grupo*, estaremos nos referindo a uma amostra de indivíduos de uma *população*. Neste trabalho, o conceito de grupo está diretamente ligado ao processo de amostragem — e estaremos sempre considerando o processo de amostragem aleatória simples. Portanto, quando falarmos em um único grupo de respondentes, nos referimos a uma amostra de indivíduos retirada de uma mesma população. Consequentemente, dois grupos – ou mais – de respondentes são dois conjuntos distintos de indivíduos, que foram amostrados de duas – ou mais – populações.

Na área de Avaliação Educacional é comum que uma população seja definida por determinadas características que podem variar, dependendo dos objetivos do estudo, e portanto, podem ou não ser relevantes para a diferenciação de populações. Por exemplo, pode-se considerar que a 5.^{a} série do Ensino Fundamental de São Paulo é a população alvo. Daí, toma-se uma única amostra dos alunos dessa população, composta de alunos do período diurno e do noturno. Nesse caso, temos então um único grupo de respondentes. Já em outro estudo, poderíamos considerar a 5.^{a} série *diurna* e a 5.^{a} série *noturna* do Ensino Fundamental de São Paulo como duas populações de interesse. Então, seriam tomadas duas amostras: uma dos alunos do período diurno e outra dos alunos do noturno. Nessa situação, teríamos dois grupos de alunos. Portanto, é pelo próprio processo de amostragem do estudo que identifica-se quantas (e quais) populações estão envolvidas.

Exemplos do que usualmente são consideradas como populações distintas são: séries distintas (3.^{a} série e 4.^{a} série); períodos distintos (diurno e noturno); uma mesma série, mas em anos distintos (3.^{a} série de 1996 e 3.^{a} série de 1997), etc.

A seguir, apresentaremos os modelos mais utilizados quando um teste é aplicado a um único grupo de respondentes.

2.2.1 Modelos para itens dicotômicos ou dicotomizados

Os modelos apresentados nesta subseção, podem ser utilizados tanto para a análise de itens de múltipla escolha dicotomizados (corrigidos como certo

ou errado) quanto para a análise de itens abertos (de resposta livre), quando avaliados de forma dicotomizada.

Na prática, os modelos logísticos para itens dicotômicos são os modelos de resposta ao item mais utilizados, sendo que há basicamente três tipos, que se diferenciam pelo número de parâmetros que utilizam para descrever o item. Eles são conhecidos como os modelos logísticos de 1, 2 e 3 parâmetros, que consideram, respectivamente:

(i) somente a dificuldade do item;

(ii) a dificuldade e a discriminação;

(iii) a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos de baixa habilidade.

Neste livro, daremos maior ênfase à explicação do modelo logístico de 3 parâmetros, uma vez que é o mais completo e portanto os outros dois podem ser facilmente obtidos a partir dele.

O modelo logístico de 3 parâmetros (ML3)

Definição

Dos modelos propostos pela TRI, o *modelo logístico unidimensional de 3 parâmetros (ML3)* é atualmente o mais utilizado e é dado por:

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

com $i = 1, 2, \cdots, I$, e $j = 1, 2, \cdots, n$, onde:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.

θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item – FRI.

b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.

a_i é o parâmetro de discriminação (ou de inclinação) do item i, com valor proporcional à inclinação da Curva Característica do Item — CCI no ponto b_i.

c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

Interpretação e representação gráfica

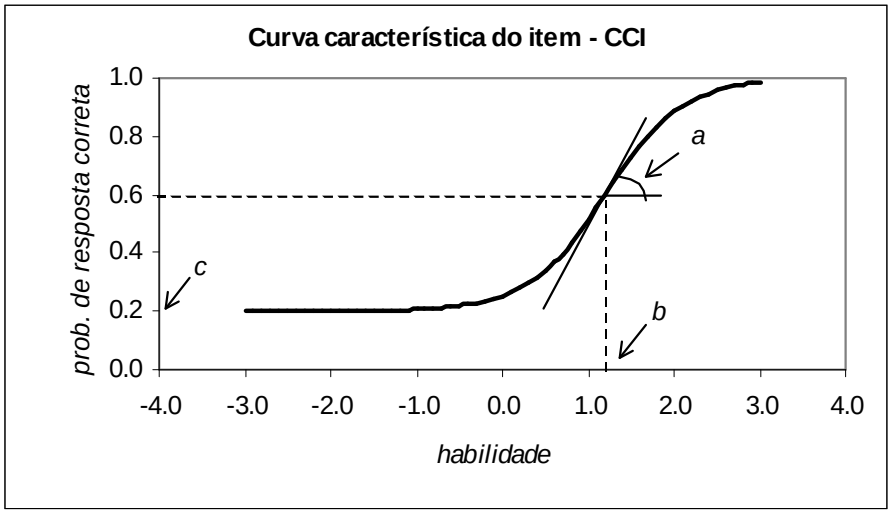
Note que $P(U_{ij} = 1|\theta_j)$ pode ser vista como a proporção de respostas corretas ao item i dentre todos os indivíduos da população com habilidade θ_j. A relação existente entre $P(U_{ij} = 1|\theta_j)$ e os parâmetros do modelo é mostrada na Figura 2.1, que é chamada de Curva Característica do Item – CCI.

O modelo proposto baseia-se no fato de que indivíduos com maior habilidade possuem maior probabilidade de acertar o item e que esta relação não é linear. De fato, pode-se perceber a partir do gráfico acima que a CCI tem forma de "S"com inclinação e deslocamento na escala de habilidade definidos pelos parâmetros do item.

A escala da habilidade é uma escala *arbitrária* onde o importante são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. O parâmetro b é medido na mesma unidade da habilidade e o parâmetro c não depende da escala, pois trata-se de uma probabilidade, e como tal, assume sempre valores entre 0 e 1.

Na realidade, o parâmetro b representa a habilidade necessária para uma

Figura 2.1 *Exemplo de uma Curva Característica do Item – CCI*



probabilidade de acerto igual a $(1 + c)/2$. Assim, quanto maior o valor de b, mais difícil é o item, e vice-versa.

O parâmetro c representa a probabilidade de um aluno com baixa habilidade responder corretamente o item e é muitas vezes referido como a probabilidade de acerto ao acaso. Então, quando não é permitido "chutar", c é igual a 0 e b representa o ponto na escala da habilidade onde a probabilidade de acertar o item é 0,5.

O parâmetro a é proporcional à derivada da tangente da curva no ponto de inflexão. Assim, itens com a negativo não são esperados sob esse modelo, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da habilidade. Baixos valores de a indicam que o item tem pouco poder de discriminação (alunos com habilidades bastante diferentes têm aproximadamente a mesma probabilidade de responder corretamente ao item) e valores muito altos indicam itens com curvas características muito "íngremes", que discriminam os alunos basicamente em dois grupos: os que possuem habilidades abaixo do valor do parâmetro b e os que possuem habilidades acima do valor do parâmetro b.

Função de Informação do Item

Uma medida bastante utilizada em conjunto com a CCI é a *função de informação do item*. Ela permite analisar quanto um item (ou teste) contém de informação para a medida de habilidade. A função de informação de um item é dada por:

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta}P_i(\theta)\right]^2}{P_i(\theta)Q_i(\theta)},$$

onde,

$I_i(\theta)$ é a "informação"fornecida pelo item i no nível de habilidade θ;

$P_i(\theta) = P(X_{ij} = 1|\theta)$ e $Q_i(\theta) = 1 - P_i(\theta)$.

No caso do modelo logístico de 3 parâmetros, a equação pode ser escrita como:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)}\left[\frac{P_i(\theta) - c_i}{1 - c_i}\right]^2.$$

Esta equação mostra a importância que têm os três parâmetros sobre o montante de informação do item. Isto é, a informação é maior:

(i) quando b_i se aproxima de θ;

(ii) quanto maior for o a_i;

(iii) e quanto mais c_i se aproximar de 0.

Função de Informação do Teste

A informação fornecida pelo teste é simplesmente a soma das informações fornecidas por cada item que compõe o mesmo:

$$I(\theta) = \sum_{i=1}^{I} I_i(\theta).$$

Outra maneira de representar esta função de informação do teste é através do erro-padrão de medida, chamado na TRI de erro-padrão de estimação, que é dado por

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}.$$

É importante notar que essas medidas de informação dependem do valor de θ. Assim, a amplitude do intervalo de confiança para θ dependerá também do seu valor.

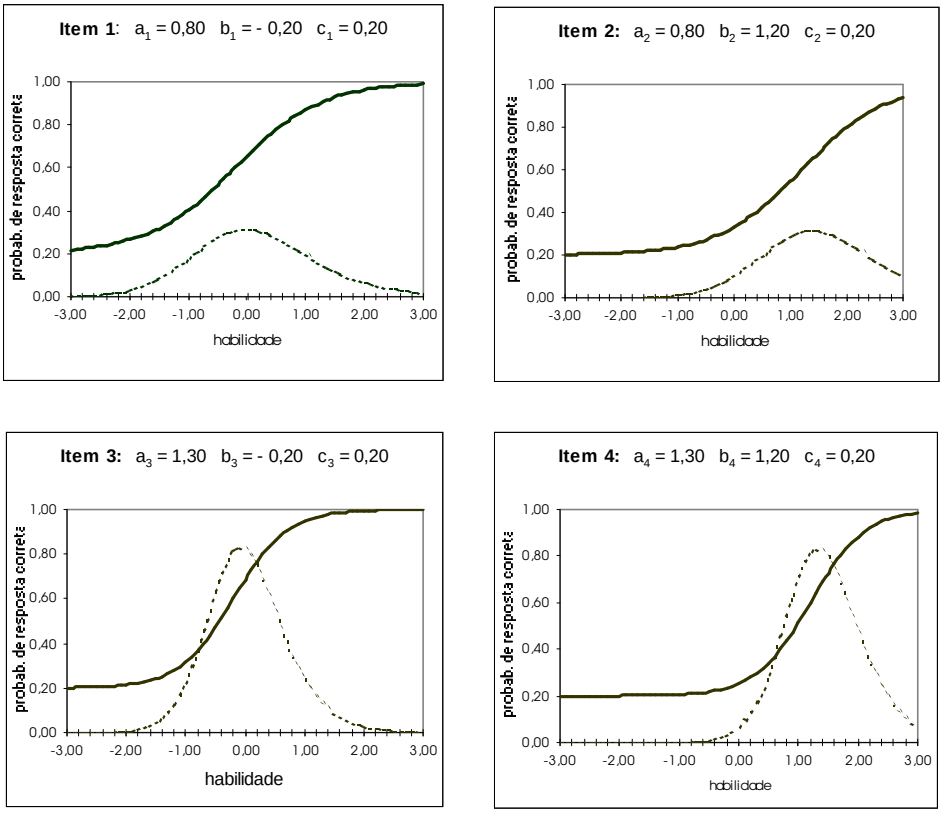
Alguns exemplos de curvas características e de curvas de informação (traçado pontilhado) de itens com diferentes combinações de valores dos parâmetros a e b são apresentados na Figura 2.2.

Comparando-se os itens 2 e 4 (e também os itens 1 e 3) pode-se perceber que os itens com maior valor do parâmetro a têm a curva característica com inclinação mais acentuada. A consequência disto é que a diferença entre as probabilidades de resposta correta de dois indivíduos com habilidades 2,00 e 1,00, por exemplo, é maior no item 4 (0,37=0,88-0,51) do que no item 2 (0,25=0,80-0,55). Em outras palavras, o item 4 é mais apropriado para discriminar estes dois indivíduos do que o item 2. Por este motivo é que o parâmetro a é denominado de *parâmetro de discriminação (ou de inclinação)* do item.

Por outro lado, comparando-se os itens 1 e 2 (e também os itens 3 e 4), pode-se perceber que os itens com maior valor do parâmetro b exigem uma habilidade maior para uma mesma probabilidade de resposta correta. Por exemplo, a habilidade requerida para uma probabilidade de resposta correta de 0,60 é igual a -0,20 no item 1 e igual a 1,20 no item 2. Isto é, o item 2 é mais difícil do que o item 1. Assim, o parâmetro b é denominado de *parâmetro de dificuldade (ou de posição)* do item.

Note que a cada item está associado um intervalo na escala de habilidade no qual o item tem maior poder de discriminação. Este intervalo é definido em torno do valor do parâmetro b e está mostrado nos gráficos pelas curvas de informação (traçados pontilhados). Deste modo, a discriminação entre bons

Figura 2.2 *Curvas características e de informação de vários itens*



alunos é feita a partir de itens considerados difíceis e não de itens considerados fáceis.

Apesar de receberem a *mesma* denominação da Teoria Clássica de Medida, o parâmetro de dificuldade do item *não* é medido por uma proporção (valor entre 0 e 1) e o parâmetro de discriminação *não* é uma correlação (valor entre -1 e 1). Na TRI, estes dois parâmetros podem, teoricamente, assumir qualquer valor real entre $-\infty$ e $+\infty$. Porém, como já foi dito, não se espera um valor negativo para o parâmetro a.

Na prática, as habilidades e os parâmetros dos itens são estimados a partir

das respostas de um grupo de indivíduos submetidos a esses itens, mas uma vez estabelecida a escala de medida da habilidade, os valores dos parâmetros dos itens não mudam, isto é, seus valores são *invariantes* a diferentes grupos de respondentes, desde que os indivíduos destes grupos tenham suas habilidades medidas na mesma escala.

A Escala de Habilidade

Diferentemente da medida escore em um teste com I questões do tipo certo/errado, que assume valores inteiros entre 0 e I, na TRI a habilidade pode teoricamente assumir qualquer valor real entre $-\infty$ e $+\infty$. Assim, precisa-se estabelecer uma origem e uma unidade de medida para a definição da escala. Esses valores são escolhidos de modo a representar, respectivamente, o valor médio e o desvio-padrão das habilidades dos indivíduos da população em estudo. Para os gráficos mostrados anteriormente, utilizou-se a escala com média igual a 0 e desvio-padrão igual a 1, que será representada por escala (0,1). Essa escala é bastante utilizada pela TRI, e nesse caso, os valores do parâmetro b variam (tipicamente) entre -2 e +2. Com relação ao parâmetro a, espera-se valores entre 0 e +2, sendo que os valores mais apropriados de a seriam aqueles maiores do que 1.

Apesar da frequente utilização da escala (0,1), em termos práticos, não faz a menor diferença estabelecer-se estes valores ou outros quaisquer. O importante são as relações de ordem existentes entre seus pontos. Por exemplo, na escala (0,1) um indivíduo com habilidade 1,20 está 1,20 desvios-padrão acima da habilidade média. Este mesmo indivíduo teria a habilidade 248, e consequentemente estaria também 1,20 desvios-padrão acima da habilidade média, se a escala utilizada para esta população fosse a escala(200;40). Isto pode ser visto a partir da transformação de escala:

$$a(\theta - b) = (a/40)[(40 \times \theta + 200) - (40 \times b + 200)] = a^*(\theta^* - b^*),$$

onde $a(\theta - b)$ é a parte do modelo probabilístico proposto envolvida na transformação. Assim, tem-se que:

1. $\theta^* = 40 \times \theta + 200$,

2. $b^* = 40 \times b + 200$,
3. $a^* = a/40$,
4. $P(U_i = 1|\theta) = P(U_i = 1|\theta^*)$.

Por exemplo, os valores dos parâmetros a e b do item 1 mostrado anteriormente, na escala (0,1) são, respectivamente, 0,80 e -0,20 e seus correspondentes na escala(200;40) são, respectivamente, 0,02 = 0,80 / 40 e 192 = 40 × (-0,20) + 200. Além disso, um indivíduo com habilidade $\theta = 1,00$ medida na escala (0,1) tem sua habilidade representada por $\theta^* = 40 \times 1{,}00 + 200 = 240$ na escala(200;40) e

$$\begin{aligned} P(U_1 = 1|\theta = 1) &= 0,20 + (1 - 0,20)\frac{1}{1 + e^{-1,7\times 0,80\times(1-(-0,20))}} \\ &= 0,20 + (1 - 0,20)\frac{1}{1 + e^{-1,7\times 0,02\times(240-192)}} \\ &= P(U_1 = 1|\theta^* = 240) = 0,87, \end{aligned}$$

ou seja, a probabilidade de um indivíduo responder corretamente a um certo item é sempre a mesma, independentemente da escala utilizada para medir a sua habilidade, ou ainda, a habilidade de um indivíduo é *invariante* à escala de medida. Assim, não faz qualquer sentido querermos analisar itens a partir dos valores de seus parâmetros a e b sem conhecer a escala na qual eles foram determinados.

Suposições do Modelo: Unidimensionalidade e Independência Local

O modelo proposto pressupõe a unidimensionalidade do teste, isto é, a homogeneidade do conjunto de itens que supostamente devem estar medindo um *único* traço latente. Em outras palavras, deve haver apenas uma habilidade responsável pela realização de todos os itens da prova. Parece claro que qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que mais de um traço latente entra na execução de qualquer tarefa. Contudo, para satisfazer o postulado da unidimensionalidade, é suficiente admitir que haja uma habilidade *dominante* (um fator dominante) responsável pelo conjunto de itens. Este fator é o que se supõe estar sendo medido pelo teste.

Tipicamente, a dimensionalidade do teste é verificada através da análise fatorial, feita a partir da matriz de correlações tetracóricas. Mislevy (1986b) discute as deficiências da aplicação deste procedimento e sugere um outro procedimento baseado no método de máxima verossimilhança.

Uma outra suposição do modelo é a chamada independência local ou independência condicional, a qual assume que para uma dada habilidade as respostas aos diferentes itens da prova são independentes. Esta suposição é fundamental para o processo de estimação dos parâmetros do modelo. Na realidade, como a unidimensionalidade implica independência local (veja Hambleton & Swaminathan (1991)), tem-se somente uma e não duas suposições a serem verificadas. Assim, itens devem ser elaborados de modo a satisfazer a suposição de unidimensionalidade.

As vantagens da utilização da TRI dependem fundamentalmente da adequação (ajuste) dos modelos e seus pressupostos. Por exemplo, somente a partir de modelos com bom ajuste é que pode-se garantir a obtenção de itens e habilidades invariantes. Uma excelente discussão das consequências da utilização de modelos inadequados aos dados e de métodos para verificação do ajuste e dos pressupostos do modelo utilizado, está apresentada no Capítulo 4 de Hambleton, Swaminathan & Rogers.

Outros modelos para itens dicotômicos

Dois outros modelos podem ser facilmente obtidos a partir do modelo logístico de 3 parâmetros. Por exemplo, quando não existe possibilidade de acerto ao acaso, pode-se considerar $c = 0$ no modelo anterior e tem-se o chamado *modelo logístico unidimensional de 2 parâmetros (ML2)*, dado por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \tag{2.2}$$

com $i = 1, 2, \cdots, I$, e $j = 1, 2, \cdots, n$.

Se além de não existir resposta ao acaso ainda tivermos todos os itens com o *mesmo* poder de discriminação, tem-se o chamado *modelo logístico unidimensional de 1 parâmetro (ML1)*, também conhecido como modelo de Rasch. Este modelo é dado por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}}, \tag{2.3}$$

com $i = 1, 2, \cdots, I,$ e $j = 1, 2, \cdots, n.$

2.2.2 Modelos para itens não dicotômicos

Aqui são incluídos os modelos tanto para a análise de itens abertos (de resposta livre) quanto para a análise de itens de múltipla escolha que são avaliados de forma graduada, ou seja, itens que são elaborados ou corrigidos de modo a ter-se uma ou mais categorias intermediárias ordenadas entre as categorias certo e errado. Nesse tipo de item não se considera somente se o indivíduo respondeu à alternativa correta ou não, mas também leva-se em conta qual foi a resposta dada por ele.

Modelo de Resposta Nominal (Nominal Categories Model)

Bock (1972) desenvolveu um modelo baseado no modelo logístico de dois parâmetros que pode ser aplicado a *todas* as categorias de resposta escolhidas em um teste com itens de múltipla escolha. O propósito deste *modelo de resposta nominal* foi maximizar a precisão da habilidade estimada usando toda a informação contida nas respostas dos indivíduos, e não apenas se o item foi respondido corretamente ou não. Bock assumiu que a probabilidade com que um indivíduo j selecionaria uma particular opção k (de m_i opções avaliáveis) do item i seria representada por:

$$P_{i,k}(\theta_j) = \frac{e^{a^+_{i,k}(\theta_j - b^+_{i,k})}}{\sum_{h=1}^{m_i} e^{a^+_{i,h}(\theta_j - b^+_{i,h})}}, \tag{2.4}$$

com $i = 1, 2, \cdots, I,$ $j = 1, 2, \cdots, n,$ e $k = 1, 2, \cdots, m_i$. Em cada θ_j, a soma das probabilidades sobre as m_i opções, $\sum_{k=1}^{m_i} P_{i,k}(\theta_j)$, é 1. As quantidades $(b^+_{i,k}; a^+_{i,k})$ são parâmetros do item i relacionados a k-ésima opção. O modelo assume que *não há nenhuma ordenação* a priori das opções de resposta.

Modelo de Resposta Gradual (Graded Response Model)

O *modelo de resposta gradual* de Samejima (1969) assume que as categorias de resposta de um item podem ser *ordenadas* entre si. Este modelo, como o modelo de Bock, tenta obter mais informação das respostas dos indivíduos do que simplesmente se eles deram respostas corretas ou incorretas.

Suponha que os escores das categorias de um item i são arranjados em ordem do menor para o maior e denotados por $k = 0, 1, \cdots, m_i$ onde $(m_i + 1)$ é o número de categorias do i-ésimo item. A probabilidade de um indivíduo j escolher *uma particular categoria ou outra mais alta* do item i pode ser dada por uma extensão do modelo logístico de 2 parâmetros:

$$P^{+}_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}}, \tag{2.5}$$

com $i = 1, 2, \cdots, I,\ j = 1, 2, \cdots, n,$ e $k = 0, 1, \cdots, m_i$, onde:

$b_{i,k}$ é o parâmetro de dificuldade da k-ésima categoria do item i.

Os demais parâmetros no modelo são análogos aos já definidos anteriormente.

No caso dos modelos para itens dicotômicos, o parâmetro de inclinação do item pode ser chamado de discriminação do item. Entretanto, no caso de modelos para itens não dicotômicos, a discriminação de uma categoria específica de resposta depende tanto do parâmetro de inclinação, comum a todas as categorias do item, quanto da distância das categorias de dificuldade adjacentes.

Cabe ressaltar que, da definição, devemos ter:

$$b_{i,1} \leq b_{i,2} \leq \ldots \leq b_{i,m_i},$$

ou seja, devemos ter necessariamente uma *ordenação* entre o nível de dificuldade das categorias de um dado item, de acordo com a classificação de seus escores.

A probabilidade de um indivíduo j receber um escore k no item i é dada então pela expressão:

$$P_{i,k}(\theta_j) = P^{+}_{i,k}(\theta_j) - P^{+}_{i,k+1}(\theta_j).$$

Samejima também define $P^+_{i,0}(\theta_j)$ e $P^+_{i,m_i+1}(\theta_j)$ de modo que:

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

e

$$P^+_{i,m_i+1}(\theta_j) = 0.$$

Portanto,

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

e

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

Então, temos que:

$$P_{i,k}(\theta_j) = \frac{1}{1+e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1+e^{-Da_i(\theta_j - b_{i,k+1})}}. \tag{2.6}$$

Note que em um item com $(m_i + 1)$ categorias, m_i valores de dificuldade necessitam ser estimados, além do parâmetro de inclinação do item. Assim, para cada item, o número de parâmetros a ser estimado será dado pelo seu número de categorias de resposta. Se, por exemplo, tivermos um teste com I itens, cada um com (m_i+1) categorias de resposta, teremos então $\left[\sum_{i=1}^{I} m_i + I\right]$ parâmetros de item a serem estimados.

Modelo de Escala Gradual (Rating Scale Model)

Um caso particular do modelo de resposta gradual de Samejima é o *modelo de escala gradual*. Analogamente ao modelo de resposta gradual, este modelo também é adequado para itens com categorias de resposta *ordenadas*. No entanto, aqui é feita uma suposição a mais: a de que os escores das categorias são *igualmente espaçados*.

Este modelo, proposto por Andrich (1978), é dado por:

$$P_{i,k}(\theta_j) = \frac{1}{1+e^{-Da_i(\theta_j - b_i + d_k)}} - \frac{1}{1+e^{-Da_i(\theta_j - b_i + d_{k+1})}}, \qquad (2.7)$$

com $i = 1, 2, \cdots, I,\ j = 1, 2, \cdots, n,$ e $k = 0, 1, \cdots, m$, onde:

b_i é agora o parâmetro de locação do item i e

d_k é o parâmetro de categoria.

Como $P^+_{i,k}(\theta_j) - P^+_{i,k+1}(\theta_j) \geq 0$, então, $d_k - d_{k+1} \geq 0$. Ou seja, devemos ter:

$$d_1 \geq d_2 \geq \cdots \geq d_m.$$

Note que a maior distinção entre o modelo de resposta gradual e o modelo de escala gradual está na hipótese de nesse último os escores das categorias de resposta devem ser equidistantes. Assim, no modelo de escala gradual o parâmetro $b_{i,k}$ é decomposto em um parâmetro b_i de locação do item e num parâmetro de categoria d_k, isto é:

$$b_{i,k} = b_i - d_k.$$

Cabe ressaltar que os parâmetros de categoria d_k não dependem do item, isto é, são *comuns* a todos os itens do teste. Logo, se os itens que compõem a prova tiverem suas próprias categorias de resposta, que podem diferir no número, então este modelo não é adequado.

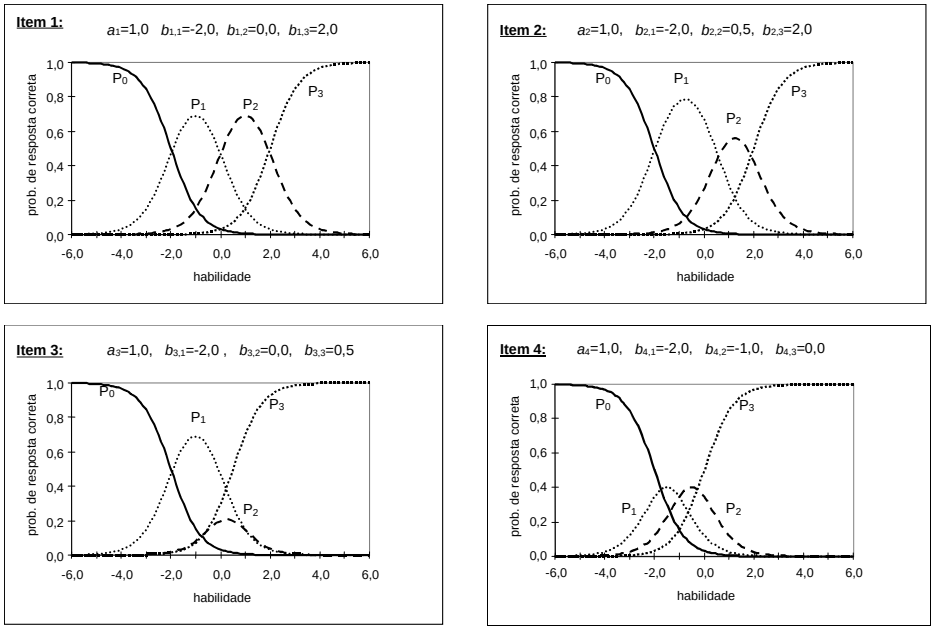
Em um teste composto por itens com $(m+1)$ categorias de resposta cada um, m parâmetros de categoria necessitam ser estimados, além dos parâmetros de inclinação e de locação de cada item. Logo, se o teste tiver I itens, teremos $[2I + m]$ parâmetros de item a serem estimados.

Na Figura 2.3 temos a representação gráfica do modelo de escala gradual e do modelo de resposta gradual para alguns itens com 4 categorias de resposta.

Em todos os itens, o parâmetro a_i foi mantido igual a 1,0. Dessa maneira, podemos verificar a importância dos parâmetros de categoria $b_{i,k}$. Os itens 1 e 4, por terem os parâmetros de categoria igualmente espaçados, podem ser representantes do modelo de escala gradual. Já o modelo de resposta gradual poderia ser representado por qualquer um dos itens acima.

Observando o item 1, podemos notar que indivíduos com habilidade até -2,0 têm maior probabilidade de responder apenas a categoria 0. Já indivíduos

Figura 2.3 *Representação gráfica dos modelos de escala gradual e de resposta gradual*



com habilidades entre -2,0 e 0,0, têm mais chance de alcançarem a categoria 1. Para habilidades entre 0,0 e 2,0, a maior probabilidade é que os indivíduos respondam até a categoria 2. Finalmente, indivíduos com habilidade acima de 2,0, devem alcançar a última categoria de resposta (que deverá representar o acerto total).

Note que do item 1 para o 2, a distância entre $b_{i,2}$ e $b_{i,3}$ tornou-se menor. A consequência disto é que aumenta a faixa de habilidade em que os indivíduos deverão responder somente até a categoria 1: de -2,0 a 0,0 no item 1 para -2,0 a 0,5 no item 2. Em outras palavras, a categoria 2 ficou mais difícil de ser alcançada, uma vez que no item 1 indivíduos com habilidades entre 0,0 e 2,0 têm *maior* probabilidade de conseguir responder à essa categoria do que indivíduos com habilidades entre 0,5 e 2,0 no item 2.

No item 3, praticamente não há chance dos indivíduos responderem até a categoria 2: indivíduos com habilidade entre -2,0 e 0,0 têm mais chance de

conseguir responder somente à categoria 1, enquanto que os indivíduos com habilidade maior do que esse valor já têm maior probabilidade de atingir a última categoria do item.

Finalmente, o item 4 é um exemplo de item onde a maioria dos indivíduos ou responde somente à primeira categoria, ou consegue chegar até a última. Apenas indivíduos com habilidades entre -2,0 e 0,0 apresentam uma chance maior de responderem somente às categorias 1 e 2.

Modelo de Crédito Parcial (Partial Credit Model)

O *modelo de crédito parcial* foi desenvolvido por Masters (1982) e é também um modelo para análise de respostas obtidas de duas ou mais categorias *ordenadas*. Nesse sentido, esse modelo é utilizado com os mesmos propósitos que outros já citados, inclusive o modelo de resposta gradual. O modelo de crédito parcial difere do gradual, entretanto, por pertencer à família de modelos de Rasch. Na verdade, o modelo de crédito parcial é uma extensão do modelo de Rasch para itens dicotômicos. Logo, todos os parâmetros no modelo são de locação, sendo que o poder de discriminação é assumido ser comum para todos os itens.

Supondo que o item i tem $(m_i + 1)$ categorias de resposta ordenáveis ($k = 0, 1, \ldots, m_i$), temos que o modelo de crédito parcial é dado por:

$$P_{i,k}(\theta_j) = \frac{\exp\left[\sum_{u=0}^{k}(\theta_j - b_{i,u})\right]}{\sum_{u=0}^{m_i}\exp\left[\sum_{v=0}^{u}(\theta_j - b_{i,v})\right]}, \tag{2.8}$$

com $i = 1, 2, \cdots, I,\ j = 1, 2, \cdots, n,\ k = 0, 1, \cdots, m_i$ e $b_{i,0} \equiv 0$, onde:

$P_{i,k}(\theta_j)$ é a probabilidade de um indivíduo com habilidade θ_j escolher a categoria k, dentre as $(m_i + 1)$ categorias do item i.

$b_{i,k}$ é o parâmetro de item que regula a probabilidade de escolher a categoria k em vez da categoria adjacente $(k - 1)$ no item i. Cada parâmetro $b_{i,k}$ corresponde ao valor de habilidade em que o indivíduo tem a mesma probabilidade de responder à categoria k e à categoria $(k - 1)$, isto é, onde $P_{i,k}(\theta_j) = P_{i,k-1}(\theta_j)$.

Assim, para itens com $(m_i + 1)$ categorias de resposta, será necessário estimar m_i parâmetros de item. Note que para itens com apenas 2 categorias de resposta, este modelo fica análogo ao modelo de Rasch para itens dicotômicos.

Modelo de Crédito Parcial Generalizado (Generalized Partial Credit Model)

O *modelo de crédito parcial generalizado — MCPG* foi formulado por Muraki (1992), que se baseou no modelo de créditos parciais de Masters, relaxando a hipótese de poder de discriminação uniforme para todos os itens. O modelo de crédito parcial generalizado é dado por:

$$P_{i,k}(\theta_j) = \frac{exp\Big[\sum_{u=0}^{k} Da_i(\theta_j - b_{i,u})\Big]}{\sum_{u=0}^{m_i} exp\Big[\sum_{v=0}^{u} Da_i(\theta_j - b_{i,v})\Big]}, \tag{2.9}$$

com $i = 1, 2, \cdots, I,\ j = 1, 2, \cdots, n,$ e $k = 0, 1, \cdots, m_i$.

Se o número de categorias de respostas é $(m_i + 1)$, somente m_i parâmetros de categoria do item podem ser identificados. Qualquer um dos $(m_i + 1)$ parâmetros de dificuldade das categorias pode ser arbitrariamente definido com qualquer valor. A razão é que o termo incluso no parâmetro é cancelado no numerador e no denominador do modelo. Em geral, definimos $b_{i,0} \equiv 0$.

Os parâmetros de categoria do item, $b_{i,k}$, são os pontos na escala de j em as curvas de $P_{i,k-1}(\theta_j)$ e $P_{i,k}(\theta_j)$ se interceptam. Essas duas funções se interceptam somente uma vez, e a intersecção pode ocorrer em qualquer ponto da escala θ_j. Então, sob a hipótese de que $a_i > 0$,

- se $\theta_j = b_{i,k}$ então $P_{i,k}(\theta_j) = P_{i,k-1}(\theta_j)$,
- se $\theta_j > b_{i,k}$ então $P_{i,k}(\theta_j) > P_{i,k-1}(\theta_j)$,
- se $\theta_j < b_{i,k}$ então $P_{i,k}(\theta_j) < P_{i,k-1}(\theta_j)$.

Da mesma maneira como no modelo de escala gradual, no MCPG o parâmetro $b_{i,k}$ pode ser decomposto como:

$$b_{i,k} = b_i - d_k.$$

Mas, é importante ressaltar que, diferentemente do modelo de escala gradual, aqui os valores de d_k *não* são necessariamente ordenados sequencialmente dentro de um item. O parâmetro d_k é interpretado como a dificuldade relativa da categoria k em comparação com as outras categorias do item ou o desvio da dificuldade de cada categoria em relação à locação do item, b_i.

Assim, em testes compostos por itens com $(m_i + 1)$ categorias de resposta, m_i parâmetros de categoria necessitam ser estimados, além dos parâmetros de inclinação e de locação de cada item. Logo, se tivermos um teste com I itens, teremos $\left[\sum_{i=1}^{I} m_i + 2I\right]$ parâmetros de item a serem estimados.

2.3 Modelos envolvendo duas ou mais populações

Alguns modelos já foram desenvolvidos para serem aplicados quando um teste envolve mais de uma população, sendo basicamente, extensões dos modelos até aqui apresentados. No entanto, um dos poucos modelos que já se encontram implementados computacionalmente e que portanto, já estão sendo utilizados na prática, quando um teste é aplicado a dois ou mais grupos de respondentes, é uma generalização dos modelos logísticos unidimensionais de 1, 2 e 3 parâmetros, que foi recentemente proposta por Bock & Zimowski (1997). O modelo é dado por:

$$P(U_{ijk} = 1|\theta_{jk}) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_{jk} - b_i)}}, \tag{2.10}$$

com $i = 1, 2, \cdots, I,\ j = 1, 2, \cdots, n_k,$ e $k = 1, 2, \cdots, K$, onde:

U_{ijk} é uma variável dicotômica que assume os valores 1, quando o indivíduo j da população k responde corretamente ao item i, ou 0 quando o indivíduo não responde corretamente ao item.

θ_{jk} representa a habilidade do j-ésimo indivíduo da população k.

$P(U_{ijk} = 1|\theta_{jk})$ é a probabilidade de um indivíduo j da população k, com habilidade θ_{jk}, responder corretamente ao item i.

Os demais parâmetros já foram descritos anteriormente.

Em geral, indivíduos pertencentes a diferentes populações não são submetidos todos aos mesmos itens. Mas, para que seja possível a comparação entre populações, é necessário que haja pelo menos alguns itens comuns entre elas. Assim, I representa o número total de itens *distintos* apresentados.

A recente implementação computacional desse modelo para mais de um grupo de respondentes foi um dos maiores avanços da TRI nos últimos anos. Através dele a comparação de indivíduos de grupos distintos, submetidos a provas diferentes mas com itens comuns, passou a ser feita de uma maneira bem mais eficiente do que era feito até então, uma vez que diminui possíveis erros de modelagem que a metodologia anterior poderia vir a ter. Algumas das questões mais importantes envolvendo a comparação de duas ou mais populações, incluindo os métodos de estimação, serão discutidas no Capítulo 5. No próximo capítulo apresentaremos os principais métodos de estimação dos parâmetros dos modelos para uma única população.

Capítulo **3**

Estimação: uma única população

3.1 Introdução

Uma das etapas mais importantes da TRI é a estimação dos parâmetros dos itens e das habilidades dos respondentes. Como foi visto no capítulo anterior, a probabilidade de uma resposta correta a um determinado item depende somente da habilidade do indivíduo e dos parâmetros que caracterizam o item. Mas, em geral, ambos são desconhecidos. Apenas as respostas dos indivíduos aos itens do teste são conhecidas.

Assim, nos modelos de resposta ao item temos um problema de estimação que envolve dois tipos de parâmetros, os parâmetro dos itens e as habilidades dos indivíduos. Então, do ponto de vista teórico, podemos dividir o problema em três situações, quando já conhecemos os parâmetros dos itens, temos apenas que estimar as habilidades; se já conhecemos as habilidades dos respondentes, estaremos interessados apenas na estimação dos parâmetros dos itens e, por fim, a situação mais comum, em que desejamos estimar os parâmetros dos itens e as habilidades dos indivíduos simultaneamente. Na TRI, o processo de estimação dos parâmetros dos itens é conhecido como calibração.

Em qualquer uma das situações citadas acima, geralmente a estimação é feita pelo Método da Máxima Verossimilhança através da aplicação de algum processo iterativo, como o algoritmo *Newton-Raphson* (ver Issac & Keller (1966), por exemplo) ou *"Scoring"de Fisher* (ver Rao (1973), por exemplo). Alguns procedimentos bayesianos também são aplicados com bastante freqüência (ver Mislevy (1986a), por exemplo).

Na situação em que desejamos estimar tanto os parâmetros dos itens, quanto as habilidades, há duas abordagens usuais: estimação conjunta, parâmetros dos itens e habilidades, ou em duas etapas, primeiro a estimação dos parâmetros dos itens e, posteriormente, das habilidades. No caso da estimação conjunta,

o número de parâmetros a serem estimados simultaneamente pode ser extremamente grande ($3I + n$, para o ML3), levando a uma enorme exigência computacional que envolve a inversão de matrizes dessa ordem. Para contornar esse problema, Birnbaum (1968) propôs um processo *vai e volta* ("back-and-forth"), que é iniciado com estimativas grosseiras das habilidades (escores padronizados, por exemplo) e envolve a estimação dos parâmetros dos itens considerando as habilidades conhecidas; após a obtenção das estimativas dos parâmetros dos itens, as estimações das habilidades são feitas considerando conhecidos os parâmetros dos itens. Esses passos são repetidos até que algum critério de parada do processo seja alcançado. A grande vantagem desse método é que ele permite, a partir da Independência Local discutida no Capítulo 2, que os itens sejam estimados individualmente, o que exige o tratamento de matrizes 3×3 para o ML3. De forma similar, a partir da independência entre as respostas oriundas de indivíduos diferentes, as habilidades também são estimadas individualmente, e com isso a exigência computacional diminui drasticamente. Entretanto, esse procedimento tem um problema sério: sabe-se que, para os parâmetros dos itens conhecidos, os *Estimadores de Máxima Verossimilhança* (EMV) das habilidades convergem (ver Sen & Singer (1993), por exemplo) para os seus verdadeiros valores quando o número de itens cresce; com as habilidades conhecidas, os EMV dos parâmetros dos itens, $\widehat{\boldsymbol{\zeta}}_i$, convergem para os seus verdadeiros valores quando o número de indivíduos cresce. Na estimação conjunta, as habilidades são denominadas de *parâmetros incidentais*, pois o número destes parâmetros (θ_j) cresce com o número de indivíduos; os parâmetros dos itens são denominados de *parâmetros estruturais*, e o número desses parâmetros não se altera quando a amostra cresce. Essas denominações são devidas a Neyman & Scott (1948), que notaram, em um contexto diferente ao da TRI, que na presença de parâmetros incidentais o EMV dos parâmetros dos itens pode ser assintoticamente viesado. Esse problema de falta de consistência dos parâmetros dos itens (ou habilidades) na presença de um número muito grande de indivíduos (ou itens) foi notado por Andersen (1973) e demonstrado para o modelo de Rasch (ML1). Porém, quando o número de itens e o número de indivíduos crescem, os EMV dos parâmetros dos itens e das habilidades podem ser não-viciados, como sugerido por Lord (1968) e demonstrado apenas para o modelo de Rasch por Haberman (1975). Resultados numéricos obtidos por Lord (1975) e Swaminathan &

Gifford (1983) reforçam a conjectura de que os EMV dos parâmetros dos itens e das habilidades são não-viciados, quando o número de itens e o número de indivíduos crescem.

O problema de possível inconsistência dos estimadores obtidos em uma etapa levou ao desenvolvimento da estimação em duas etapas por Bock & Lieberman (1970). Este método baseia-se na existência de uma distribuição (latente) associada à habilidade dos indivíduos da população em estudo Π (ver Andersen (1980) para maiores detalhes). Isso possibilita que a estimação dos itens seja feita pelo *Método da Máxima Verossimilhança Marginal*, ou seja, considerando uma determinada distribuição para a habilidade dos indivíduos de Π, cuja função densidade de probabilidade (fdp) é $g(\theta|\boldsymbol{\eta})$, onde $\boldsymbol{\eta}$ é o conjunto de parâmetros associados à Π, e integrando a função de verossimilhança com relação a θ. Após a estimação dos parâmetros dos itens, as habilidades são estimadas individualmente por máxima verossimilhança ou pela moda ou média da distribuição condicional de θ_j dado $\boldsymbol{u}_{j\cdot} = (u_{j1}, \cdots, u_{jI})$, o vetor de respostas do indivíduo j, $j = 1, \cdots, n$, com $\boldsymbol{\zeta}_i = (a_i, b_i, c_i)$, o vetor de parâmetros do item i, $i = 1, \cdots, I$, conhecidos. Embora este método tenha a vantagem de envolver, na primeira etapa, apenas a estimação dos parâmetros dos itens, a estimação é feita através de aplicação de métodos numéricos que dependem das derivadas segundas da log-verossimilhança com relação a $\boldsymbol{\zeta}_i$ e $\boldsymbol{\zeta}_k$, $i, k = 1, \cdots, I$, que podem ser não nulas para $i \neq k$. Com isso, há a necessidade da inversão de matrizes de ordem $3I \times 3I$ para o ML3, o que ainda pode ser bastante exigente do ponto de vista computacional. Para contornar esse problema, Bock & Aitkin (1981) fizeram uma modificação no modelo de Bock & Lieberman adicionando a suposição de independência entre os itens, de forma que as derivadas segundas citadas acima para $i \neq k$ sejam nulas. Com isso, a matriz $3I \times 3I$ (no ML3) de derivadas segundas torna-se bloco-diagonal, o que possibilita que os (parâmetros dos) itens sejam estimados individualmente. Adicionalmente, Bock & Aitkin propõem que a obtenção das estimativas de máxima verossimilhança seja feita com a aplicação do algoritmo EM introduzido por Dempster, Laird & Rubin (1977).

Embora existam outras propostas de estimação para os parâmetros dos itens e habilidades, as citadas acima podem ser consideradas as mais importantes e, portanto, serão exploradas nesse texto. Na Seção 3.2 consideraremos o caso da estimação dos parâmetros dos itens quando as habilidades são conhecidas.

Na Seção 3.3 consideraremos a situação contrária: estimação das habilidades com os parâmetros dos itens conhecidos. Em complemento a essas duas seções, na Seção 3.4, trataremos da estimação conjunta: parâmetros dos itens e habilidades, em uma etapa. Na Seção 3.5 também consideraremos a estimação conjunta dos parâmetros dos itens e habilidades, mas agora em duas etapas através da máxima verossimilhança marginal. Na Seção 3.6 complementaremos a etapa de estimação considerando a abordagem bayesiana, tanto dos parâmetros dos itens quanto das habilidades. Recomenda-se a leitura de Baker (1992) para maiores detalhes das idéias e resultados que serão apresentados nesse capítulo.

Em todos os métodos de estimação descritos a seguir, algumas notações e suposições serão necessárias para o desenvolvimento do modelo. Em particular, sejam θ_j a habilidade e U_{ji} a variável aleatória que representa a resposta (binária) do indivíduo j ao item i, com

$$U_{ji} = \begin{cases} 1, & \text{resposta correta,} \\ 0, & \text{resposta incorreta.} \end{cases}$$

Sejam $\boldsymbol{U}_{j\cdot} = (U_{j1}, U_{j2}, \cdots, U_{jI})$ o vetor aleatório de respostas do indivíduo j e $\boldsymbol{U}_{\cdot\cdot} = (\boldsymbol{U}_{1\cdot}, \cdots, \boldsymbol{U}_{n\cdot})$ o conjunto integral de respostas. De forma similar, representaremos as observações por $u_{ji}, \boldsymbol{u}_{j\cdot}$ e $\boldsymbol{u}_{\cdot\cdot}$. Ainda, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ representará o vetor de habilidades dos n indivíduos e $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_I)$ o conjunto de parâmetros dos itens.

As duas principais suposições que usaremos em todo o restante deste texto, são as seguintes:

(S1) as respostas oriundas de indivíduos diferentes são independentes,

(S2) os itens são respondidos de forma independente por cada indivíduo (Independência Local), fixada sua habilidade.

A suposição (S2) necessita de uma discussão um pouco mais detalhada: ela garante que, para cada valor de θ, se tomarmos um conjunto de indivíduos com habilidade θ, as covariâncias entre as respostas para cada par de itens serão nulas. Entretanto, se for considerado um conjunto de indivíduos com habilidades variadas, estas covariâncias não necessariamente serão nulas. Na verdade, elas serão positivas (ver Lord & Novick (1968, pág. 361)).

Quando necessárias, outras suposições serão adotadas. Em algumas situações usaremos notações simplificadas. Por exemplo, as probabilidades $P(U_{ji} = u_{ji}|.)$ poderão ser representadas por $P(u_{ji}|.)$; o mesmo valendo para os vetores de observações. Poderemos, ainda, usar algumas expressões simplificadas, tais como "estimação dos itens" ao invés de "estimação dos parâmetros dos itens". As demonstrações dos principais resultados apresentados nesse capítulo poderão ser encontradas no Apêndice A.

3.2 Estimação dos parâmetros dos itens

Nesta seção trataremos da estimação dos parâmetros dos itens pelo método da máxima verossimilhança, quando as habilidades são conhecidas. Embora qualquer modelo descrito no Capítulo 2 possa ser adotado para fins de aplicação, o ML3 tem sido amplamente empregado e, por isso, o usaremos para efeito de exemplificação. Os modelos ML1 e ML2 são casos particulares do ML3; a escolha desse último leva a resultados que servem para os dois primeiros.

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

$$\begin{aligned} L(\boldsymbol{\zeta}) &= \prod_{j=1}^{n} P(\boldsymbol{U}_{j.} = \boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta}) \\ &= \prod_{j=1}^{n}\prod_{i=1}^{I} P(U_{ji} = u_{ji}|\theta_j, \boldsymbol{\zeta}_i), \end{aligned}$$

onde na última igualdade usamos que a distribuição de U_{ji} só depende de $\boldsymbol{\zeta}$ através de $\boldsymbol{\zeta}_i$. Usando a notação $P_{ji} = P(U_{ji} = 1|\theta_j, \boldsymbol{\zeta}_i)$ e $Q_{ji} = 1 - P_{ji}$, temos que

$$\begin{aligned} P(U_{ji} = u_{ji}|\theta_j, \boldsymbol{\zeta}_i) &= P(U_{ji} = 1|\theta_j, \boldsymbol{\zeta}_i)^{u_{ji}} P(U_{ji} = 0|\theta_j, \boldsymbol{\zeta}_i)^{1-u_{ji}} \\ &= P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}}. \end{aligned} \tag{3.1}$$

Portanto,

$$L(\boldsymbol{\zeta}) = \prod_{j=1}^{n}\prod_{i=1}^{I} P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}}. \tag{3.2}$$

Segue que a log-verossimilhança pode ser escrita como

$$\log L(\boldsymbol{\zeta}) \;=\; \sum_{j=1}^{n}\sum_{i=1}^{I}\{u_{ji}\log P_{ji} + (1-u_{ji})\log Q_{ji}\}. \tag{3.3}$$

Os Estimadores de Máxima verossimilhança (EMV) de $\boldsymbol{\zeta}_i,\ i = 1, \cdots, I$, são os valores que maximizam a verossimilhança, ou equivalente, são as soluções da equação

$$\frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} = \mathbf{0}, \quad i = 1, \cdots, I.$$

Notemos que

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} &= \sum_{j=1}^{n}\left\{u_{ji}\frac{\partial(\log P_{ji})}{\partial \boldsymbol{\zeta}_i} + (1-u_{ji})\frac{\partial(\log Q_{ji})}{\partial \boldsymbol{\zeta}_i}\right\} \\
&= \sum_{j=1}^{n}\left\{u_{ji}\frac{1}{P_{ji}}\left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i}\right) - (1-u_{ji})\frac{1}{Q_{ji}}\left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i}\right)\right\} \\
&= \sum_{j=1}^{n}\left\{u_{ji}\frac{1}{P_{ji}} - (1-u_{ji})\frac{1}{Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i}\right) \\
&= \sum_{j=1}^{n}\left\{\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i}\right).
\end{aligned} \tag{3.4}$$

Por conveniência, consideremos a seguinte ponderação:

$$W_{ji} = \frac{P_{ji}^{*}Q_{ji}^{*}}{P_{ji}Q_{ji}}, \tag{3.5}$$

onde

$$P_{ji}^{*} = \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1} \quad \text{e} \quad Q_{ji}^{*} = 1 - P_{ji}^{*}. \tag{3.6}$$

Com isso, podemos reescrever a Equação (3.4) como

$$\frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} = \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^{*} Q_{ji}^{*}} \right\} \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right). \tag{3.7}$$

Para obter as equações de estimação, precisaremos das seguintes expressões:

$$\frac{\partial P_{ji}}{\partial a_i} = D(1 - c_i)(\theta_j - b_i) P_{ji}^{*} Q_{ji}^{*}, \tag{3.8}$$

$$\frac{\partial P_{ji}}{\partial b_i} = -Da_i(1 - c_i) P_{ji}^{*} Q_{ji}^{*}, \tag{3.9}$$

$$\frac{\partial P_{ji}}{\partial c_i} = Q_{ji}^{*}. \tag{3.10}$$

Para o parâmetro de discriminação, temos de (3.7) e (3.8) que

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta})}{\partial a_i} &= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji}}{\partial a_i} \right) \frac{W_{ji}}{P_{ji}^{*} Q_{ji}^{*}} \right\} \\ &= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) D(1 - c_i)(\theta_j - b_i) P_{ji}^{*} Q_{ji}^{*} \frac{W_{ji}}{P_{ji}^{*} Q_{ji}^{*}} \right\} \\ &= D(1 - c_i) \sum_{j=1}^{n} (u_{ji} - P_{ji})(\theta_j - b_i) W_{ji}. \end{aligned} \tag{3.11}$$

Para o parâmetro de dificuldade, temos de (3.7) e (3.9) que

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta})}{\partial b_i} &= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji}}{\partial b_i} \right) \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji})(-1) D a_i (1 - c_i) P^*_{ji} Q^*_{ji} \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= -D a_i (1 - c_i) \sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji}. \qquad (3.12)
\end{aligned}$$

Para o parâmetro de acerto ao acaso, temos de (3.7) e (3.10) que

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta})}{\partial c_i} &= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji}}{\partial c_i} \right) \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) Q^*_{ji} \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P^*_{ji}} \right\}. \qquad (3.13)
\end{aligned}$$

Em resumo, as equações de estimação para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i \quad : \quad D(1 - c_i) \sum_{j=1}^{n} (u_{ji} - P_{ji})(\theta_j - b_i) W_{ji} = 0, \qquad (3.14)$$

$$b_i \quad : \quad -D a_i (1 - c_i) \sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji} = 0, \qquad (3.15)$$

$$c_i \quad : \quad \sum_{j=1}^{n} (u_{ji} - P_{ji}) \frac{W_{ji}}{P^*_{ji}} = 0. \qquad (3.16)$$

Embora as constantes antepostas aos somatórios em (3.14) e (3.15) possam (em princípio) ser eliminadas nas apresentação das referidas equações, vamos mantê-las por todo o restante do texto.

Agrupamento das habilidades

Um procedimento alternativo de estimação é considerar as habilidades agrupadas em q categorias. Isso é possível porque estamos considerando as habilidades conhecidas, logo podemos agrupá-las definindo um conjunto de q intervalos cujos valores centrais (ou alguma medida central dessas habilidades) sejam denotados por $\overline{\theta}_k,\ k = 1, \cdots, q$. Para fins de desenvolvimento, podemos supor que todos os indivíduos pertencentes à categoria k têm habilidade $\overline{\theta}_k$, o que pode reduzir bastante a exigência computacional tornando este método mais atrativo.

De forma geral, consideremos que q grupos de $f_{ki},\ k = 1, \cdots, q$, indivíduos com habilidades conhecidas $\overline{\theta}_k$ são selecionados ao acaso da população Π em estudo para responder ao item i. Seja r_{ki} o número de indivíduos do grupo k que responderam corretamente ao item i. Vale notar que em algumas situações os mesmos grupos de indivíduos responderão a todos os itens, e portanto poderemos representar as quantidades f_{ki} e r_{ki} por f_k e r_k, respectivamente. Ocorre que pela independência local, podemos tratar da estimação de cada item individualmente e, por isso, é conveniente usar um índice relativo ao item a ser estimado. Entretanto, o motivo principal para o uso desta notação está no fato de que, na prática, é comum que alguns indivíduos não respondam (ou anulem de outra forma) alguns itens. Isso possibilita que um grupo com n_k indivíduos apresente n_{ki} respostas ao item i e n_{kl} ao item l com $n_{ki} \neq n_{kl}$. Dessa forma, mesmo considerando (em princípio) que todos os indivíduos respondam a todos os itens, para tornar o tratamento mais geral adotaremos o índice i nas quantidadas f_{ki} e r_{ki}.

Pela independência entre as respostas dos diferentes indivíduos, podemos assumir que $r_{ki},\ k = 1, \cdots, q$, tem distribuição *Binomial* com parâmetros f_{ki} e P_{ki}, onde P_{ki} representa o ML3, com θ_j substituída por $\overline{\theta}_k$. De acordo com isso, a verossimilhança será

$$L(\boldsymbol{\zeta}) = \prod_{k=1}^{q} \prod_{i=1}^{I} \left\{ \binom{f_{ki}}{r_{ki}} P_{ki}^{r_{ki}} Q_{ki}^{f_{ki}-r_{ki}} \right\},$$

e a log-verossimilhança,

$$\log L(\boldsymbol{\zeta}) = C + \sum_{k=1}^{q}\sum_{i=1}^{I}\left\{r_{ki}\log P_{ki} + (f_{ki}-r_{ki})\log Q_{ki}\right\}, \tag{3.17}$$

onde $C = \sum_{k=1}^{q}\sum_{i=1}^{I}\log\binom{f_{ki}}{r_{ki}}$ é constante com relação a $\boldsymbol{\zeta}$. Tomando a derivada de (3.17) com relação a $\boldsymbol{\zeta}_i$, teremos

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} &= \sum_{k=1}^{q}\left\{r_{ki}\frac{1}{P_{ki}}\left(\frac{\partial P_{ki}}{\partial \boldsymbol{\zeta}_i}\right) + (f_{ki}-r_{ki})\frac{1}{Q_{ki}}\left(\frac{\partial Q_{ki}}{\partial \boldsymbol{\zeta}_i}\right)\right\} \\
&= \sum_{k=1}^{q}\frac{1}{P_{ki}Q_{ki}}(r_{ki}-f_{ki}P_{ki})\left(\frac{\partial P_{ki}}{\partial \boldsymbol{\zeta}_i}\right) \\
&= \sum_{k=1}^{q}(r_{ki}-f_{ki}P_{ki})\frac{W_{ki}}{P^*_{ki}Q^*_{ki}}\left(\frac{\partial P_{ki}}{\partial \boldsymbol{\zeta}_i}\right),
\end{aligned}$$

onde a última igualdade é devida a (3.5). Usando as expressões (3.8) a (3.10), temos que as equações de estimação para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i \quad : \quad D(1-c_i)\sum_{k=1}^{q}(r_{ki}-f_{ki}P_{ki})(\overline{\theta}_k - b_i)W_{ki} = 0, \tag{3.18}$$

$$b_i \quad : \quad -Da_i(1-c_i)\sum_{k=1}^{q}(r_{ki}-f_{ki}P_{ki})W_{ki} = 0, \tag{3.19}$$

$$c_i \quad : \quad \sum_{k=1}^{q}(r_{ki}-f_{ki}P_{ki})\frac{W_{ki}}{P^*_{ki}} = 0. \tag{3.20}$$

Estas equações, bem como (3.14) a (3.16), não possuem solução explícita e por isso precisaremos de algum método iterativo para a obtenção das estimativas de máxima verossimilhança dos parâmetros dos itens. A seguir, damos uma breve descrição do algoritmo Newton-Raphson e do método "Scoring"de Fisher.

3.2.1 Aplicação do algoritmo Newton-Raphson

Seja $l(\boldsymbol{\zeta}) = \log L(\boldsymbol{\zeta})$ a log-verossimilhança, onde $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_I)$, com $\boldsymbol{\zeta}_i = (a_i, b_i, c_i)'$. Se valores iniciais $\widehat{\boldsymbol{\zeta}}_i^{(0)} = (a_i^{(0)}, b_i^{(0)}, c_i^{(0)})'$ podem ser encontrados para $\boldsymbol{\zeta}_i$, então uma estimativa atualizada será $\widehat{\boldsymbol{\zeta}}_i^{(1)} = \widehat{\boldsymbol{\zeta}}_i^{(0)} + \Delta\widehat{\boldsymbol{\zeta}}_i^{(0)}$, ou seja,

$$\begin{aligned} \hat{a}_i^{(1)} &= \hat{a}_i^{(0)} + \Delta\hat{a}_i^{(0)}, \\ \hat{b}_i^{(1)} &= \hat{b}_i^{(0)} + \Delta\hat{b}_i^{(0)}, \\ \hat{c}_i^{(1)} &= \hat{c}_i^{(0)} + \Delta\hat{c}_i^{(0)}, \end{aligned} \tag{3.21}$$

onde $\Delta\hat{a}_i^{(0)}$, $\Delta\hat{b}_i^{(0)}$ e $\Delta\hat{c}_i^{(0)}$ são erros de aproximação. Usando a expansão em série de Taylor de $\partial l(\boldsymbol{\zeta})/\partial\boldsymbol{\zeta}_i$ em torno de $\widehat{\boldsymbol{\zeta}}_i^{(0)}$, teremos

$$\begin{aligned} \frac{\partial l(\boldsymbol{\zeta})}{\partial a_i} &= \frac{\partial l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i} + \Delta\hat{a}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i^2} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i \partial b_i} + \Delta\hat{c}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i \partial c_i} + R_{a_i}(\widehat{\boldsymbol{\zeta}}_i^{(0)}), \\ \frac{\partial l(\boldsymbol{\zeta})}{\partial b_i} &= \frac{\partial l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i^2} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i \partial a_i} + \Delta\hat{c}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i \partial c_i} + R_{b_i}(\widehat{\boldsymbol{\zeta}}_i^{(0)}), \\ \frac{\partial l(\boldsymbol{\zeta})}{\partial c_i} &= \frac{\partial l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i} + \Delta\hat{a}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i^2} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i \partial a_i} + \Delta\hat{c}_i^{(0)}\frac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i \partial b_i} + R_{c_i}(\widehat{\boldsymbol{\zeta}}_i^{(0)}), \end{aligned}$$

onde $\partial l(\widehat{\boldsymbol{\zeta}}_i)/\partial\alpha_i$ representa a função $\partial l(\boldsymbol{\zeta}_i)/\partial\alpha_i$ avaliada no ponto $\boldsymbol{\zeta}_i = \widehat{\boldsymbol{\zeta}}_i$. Nessas expressões estamos usando que $\partial l(\boldsymbol{\zeta})/\partial\boldsymbol{\zeta}_i$ é função apenas de $\boldsymbol{\zeta}_i$, não dependendo de $\boldsymbol{\zeta}_l$ para $l \neq i$. Por isso, poderemos representá-la de forma simplificada por $\partial l(\boldsymbol{\zeta}_i)/\partial\boldsymbol{\zeta}_i$. Fazendo

$$\frac{\partial l(\boldsymbol{\zeta}_i)}{\partial a_i} = \frac{\partial l(\boldsymbol{\zeta}_i)}{\partial b_i} = \frac{\partial l(\boldsymbol{\zeta}_i)}{\partial c_i} = 0,$$

usando a notação

$$
\begin{array}{llll}
L_1 = \dfrac{\partial l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i} & L_{11} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i^2} & L_{12} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i \partial b_i} & L_{13} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial a_i \partial c_i}, \\
L_2 = \dfrac{\partial l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i} & L_{21} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i \partial a_i} & L_{22} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i^2} & L_{23} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial b_i \partial c_i}, \\
L_3 = \dfrac{\partial l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i} & L_{31} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i \partial a_i} & L_{32} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i \partial b_i} & L_{33} = \dfrac{\partial^2 l(\widehat{\boldsymbol{\zeta}}_i^{(0)})}{\partial c_i^2},
\end{array}
$$

e desprezando os restos $R_{a_i}(\widehat{\boldsymbol{\zeta}}_i^{(0)})$, $R_{b_i}(\widehat{\boldsymbol{\zeta}}_i^{(0)})$ $R_{c_i}(\widehat{\boldsymbol{\zeta}}_i^{(0)})$, teremos

$$
\begin{aligned}
0 &= L_1 + L_{11}\Delta\hat{a}_i^{(0)} + L_{12}\Delta\hat{b}_i^{(0)} + L_{13}\Delta\hat{c}_i^{(0)}, \\
0 &= L_2 + L_{12}\Delta\hat{a}_i^{(0)} + L_{22}\Delta\hat{b}_i^{(0)} + L_{23}\Delta\hat{c}_i^{(0)}, \\
0 &= L_3 + L_{13}\Delta\hat{a}_i^{(0)} + L_{23}\Delta\hat{b}_i^{(0)} + L_{33}\Delta\hat{c}_i^{(0)}.
\end{aligned}
$$

Colocando o resultado em forma matricial, teremos

$$
-\begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} \Delta\hat{a}_i^{(0)} \\ \Delta\hat{b}_i^{(0)} \\ \Delta\hat{c}_i^{(0)} \end{pmatrix}.
$$

Resolvendo o sistema para $\Delta\widehat{\boldsymbol{\zeta}}_i^{(0)}$, teremos

$$
\begin{pmatrix} \Delta\hat{a}_i^{(0)} \\ \Delta\hat{b}_i^{(0)} \\ \Delta\hat{c}_i^{(0)} \end{pmatrix} = -\begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix}^{-1} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix},
$$

e finalmente, por (3.21)

$$
\begin{pmatrix} \hat{a}_i^{(1)} \\ \hat{b}_i^{(1)} \\ \hat{c}_i^{(1)} \end{pmatrix} = \begin{pmatrix} \hat{a}_i^{(0)} \\ \hat{b}_i^{(0)} \\ \hat{c}_i^{(0)} \end{pmatrix} - \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix}^{-1} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix}.
$$

Após obtido $\widehat{\boldsymbol{\zeta}}_i^{(1)}$, este é considerado um novo ponto inicial para a obtenção de $\widehat{\boldsymbol{\zeta}}_i^{(2)}$, e assim por diante. Este processo é repetido até que algum critério de parada seja alcançado. Por exemplo, até que $\Delta\widehat{\boldsymbol{\zeta}}_i^{(t)} = \widehat{\boldsymbol{\zeta}}_i^{(t)} - \widehat{\boldsymbol{\zeta}}_i^{(t-1)}$ seja suficientemente pequeno ou que um número pré-definido, t_{max}, de iterações seja cumprido.

As expressões $L_k,\ k = 1, 2, 3$ são dadas por (3.11) a (3.13), respectivamente e as expressões $L_{kl},\ k, l = 1, 2, 3$, são obtidas de

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} &= \sum_{j=1}^{n} \left\{ \left[\frac{\partial}{\partial \boldsymbol{\zeta}_i} \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) \right] \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right)' + \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) \left(\frac{\partial^2 P_{ji}}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \right) \right\} \\
&= \sum_{j=1}^{n} \left\{ \left[\frac{\partial v_{ji}}{\partial \boldsymbol{\zeta}_i} \right] \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right)' + v_{ji} \left(\frac{\partial^2 P_{ji}}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \right) \right\}, \qquad (3.22)
\end{aligned}$$

onde

$$v_{ji} = \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \qquad (3.23)$$

e

$$\begin{aligned}
\frac{\partial v_{ji}}{\partial \boldsymbol{\zeta}_i} &= \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) = \\
&= \frac{1}{(P_{ji} Q_{ji})^2} \left\{ -P_{ji} Q_{ji} \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right) - (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji} Q_{ji}}{\partial \boldsymbol{\zeta}_i} \right) \right\} \\
&= \frac{-1}{(P_{ji} Q_{ji})^2} \left\{ P_{ji} Q_{ji} \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right) + (u_{ji} - P_{ji}) \left[\left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right) - 2 P_{ji} \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right) \right] \right\} \\
&= \frac{-1}{(P_{ji} Q_{ji})^2} \{ P_{ji} Q_{ji} + (u_{ji} - P_{ji})(1 - 2P_{ji}) \} \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right) \\
&= \frac{-1}{(P_{ji} Q_{ji})^2} (u_{ji} - P_{ji})^2 \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right), \\
&= -v_{ji}^2 \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right). \qquad (3.24)
\end{aligned}$$

A última igualdade segue do fato que $u_{ji} = u_{ji}^2$.

Considerando $\widehat{\boldsymbol{\zeta}}_i^{(t)}$ a estimativa de $\boldsymbol{\zeta}_i$ na iteração t, então na iteração $t+1$ do algoritmo Newton-Raphson teremos que

$$\widehat{\boldsymbol{\zeta}}_i^{(t+1)} = \widehat{\boldsymbol{\zeta}}_i^{(t)} - [\boldsymbol{H}(\widehat{\boldsymbol{\zeta}}_i^{(t)})]^{-1}\boldsymbol{h}(\widehat{\boldsymbol{\zeta}}_i^{(t)}). \tag{3.25}$$

onde, ver Apêndice A.1 para as demonstrações dos resultados,

$$\begin{aligned}
\boldsymbol{h}(\boldsymbol{\zeta}_i) &\equiv \frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} \\
&= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^* Q_{ji}^*} \right\} (P_{ji}^* Q_{ji}^*) \boldsymbol{h}_{ji} \\
&= \sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji} \boldsymbol{h}_{ji}.
\end{aligned} \tag{3.26}$$

e

$$\begin{aligned}
\boldsymbol{H}(\boldsymbol{\zeta}_i) &\equiv \frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \\
&= \sum_{j=1}^{n} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) (P_{ji}^* Q_{ji}^*) \boldsymbol{H}_{ji} - \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right)^2 (P_{ji}^* Q_{ji}^*)^2 \boldsymbol{h}_{ji} \boldsymbol{h}_{ji}' \right\} \\
&= \sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji} \left\{ \boldsymbol{H}_{ji} - (u_{ji} - P_{ji}) W_{ji} \boldsymbol{h}_{ji} \boldsymbol{h}_{ji}' \right\}.
\end{aligned} \tag{3.27}$$

com

$$\boldsymbol{h}_{ji} = (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i}\right) = \begin{pmatrix} D(1-c_i)(\theta_j - b_i) \\ -Da_i(1-c_i) \\ \frac{1}{P^*_{ji}} \end{pmatrix},$$

e

$$\begin{aligned} \boldsymbol{H}_{ji} &= (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial^2 P_{ji}}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}\right) \\ &= \begin{pmatrix} D^2(1-c_i)(\theta_j - b_i)^2(1-2P^*_{ji}) & . & . \\ -D(1-c_i)\{1 + Da_i(\theta_j - b_i)(1-2P^*_{ji})\} & D^2a_i^2(1-c_i)(1-2P^*_{ji}) & . \\ -D(\theta_j - b_i) & Da_i & 0 \end{pmatrix}. \end{aligned}$$

Para a abordagem utilizando as habilidades agrupadas em q categorias, as expressões para (3.26) e (3.27) são

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{\zeta}_i) &= \sum_{k=1}^{q}(r_{ki} - f_{ki}P_{ki})W_{ki}\boldsymbol{h}_{ki}, \\ \boldsymbol{H}(\boldsymbol{\zeta}_i) &= \sum_{k=1}^{q}(r_{ki} - f_{ki}P_{ki})W_{ki}\left\{\boldsymbol{H}_{ki} - (r_{ki} - f_{ki}P_{ki})W_{ki}\boldsymbol{h}_{ki}\boldsymbol{h}_{ki}'\right\}. \end{aligned}$$

3.2.2 Aplicação do método "Scoring"de Fisher

Para aplicação do método "Scoring"de Fisher, devemos substituir os componentes da matriz de derivadas segundas usadas no processo iterativo de Newton-Raphson pelos seus valores esperados. Notando que a variável U_{ji} só pode assumir dois valores: 1, com probabilidade P_{ji} e 0 com probabilidade Q_{ji}, então U_{ji} tem distribuição $Bernoulli(P_{ji})$. Segue que $E(U_{ji}) = P_{ji}$ e $E(U_{ji} - P_{ji})^2 = Var(U_{ji}) = P_{ji}Q_{ji}$. Logo, de (3.27), temos que

$$\begin{aligned}
\boldsymbol{\Delta}(\boldsymbol{\zeta}_i) &\equiv E(\boldsymbol{H}(\boldsymbol{\zeta}_i)) \\
&= \sum_{j=1}^{N} \left\{ E(U_{ji} - P_{ji}) W_{ji} \boldsymbol{H}_{ji} - E(U_{ji} - P_{ji})^2 W_{ji}^2 \boldsymbol{h}_{ji} \boldsymbol{h}'_{ji} \right\} \\
&= \sum_{j=1}^{N} \left\{ -P_{ji} Q_{ji} W_{ji}^2 \boldsymbol{h}_{ji} \boldsymbol{h}'_{ji} \right\} \\
&= -\sum_{j=1}^{N} \left\{ P_{ji}^* Q_{ji}^* W_{ji} \boldsymbol{h}_{ji} \boldsymbol{h}'_{ji} \right\}. \qquad (3.28)
\end{aligned}$$

Para as habilidades agrupadas, a expressão acima fica

$$\boldsymbol{\Delta}(\boldsymbol{\zeta}_i) = -\sum_{k=1}^{q} \left\{ P_{ki}^* Q_{ki}^* W_{ki} \boldsymbol{h}_{ki} \boldsymbol{h}'_{ki} \right\}.$$

A expressão para estimativa de $\boldsymbol{\zeta}_i$ na iteração $t+1$ será

$$\widehat{\boldsymbol{\zeta}}_i^{(t+1)} = \widehat{\boldsymbol{\zeta}}_i^{(t)} - [\boldsymbol{\Delta}(\widehat{\boldsymbol{\zeta}}_i^{(t)})]^{-1} \boldsymbol{h}(\widehat{\boldsymbol{\zeta}}_i^{(t)}).$$

3.2.3 Erro-padrão

Os estimadores de máxima verossimilhança gozam de propriedades assintóticas conhecidas, tais como vício nulo e eficiência. Sob algumas condições de regularidade (ver Sen & Singer (1993), por exemplo) a distribuição assintótica do estimador de máxima verossimilhança, $\widehat{\boldsymbol{\zeta}}_i$, é normal com vetor de média $\boldsymbol{\zeta}_i$ e matriz de covariâncias dada pela inversa da matriz de informação

$$\boldsymbol{I}(\boldsymbol{\zeta}_i) = -E\left(\frac{\partial^2 \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}'_i} \right) = -\boldsymbol{\Delta}(\boldsymbol{\zeta}_i), \qquad (3.29)$$

onde $\boldsymbol{\Delta}(\boldsymbol{\zeta}_i)$ é obtida de (3.28). As raízes quadradas dos elementos diagonais de $[\boldsymbol{I}(\boldsymbol{\zeta}_i)]^{-1}$ fornecem os erros-padrão dos estimadores $\widehat{a}_i$, $\widehat{b}_i$ e $\widehat{c}_i$.

3.2.4 Escore nulo ou perfeito

Alguns problemas ocorrem na estimação por máxima verossimilhança. Se o item i é respondido incorretamente por todos os indivíduos, isto é, $u_{ji} = 0,\ j = 1, \cdots, n$, então a verossimilhança (3.2) resume-se a $L(\boldsymbol{\zeta}) = \prod_{j=1}^{n} Q_{ji}$. Considerando os valores a_i, c_i e θ_j fixos, temos que mudanças no valor de b_i apenas transladam Q_{ji}, sem alterar seus valores máximo e mínimo. Com isso, fixando $a_i, c_i,\ i = 1, \cdots, I,\ b_l,\ l \neq i$ e θ_j, o valor que maximiza a verossimilhança (EMV) será $b_i = -\infty$. Por outro lado, se o item i é respondido corretamente por todos os indivíduos, isto é, $u_{ji} = 1$, então a verossimilhança (3.2) resume-se a $L(\boldsymbol{\zeta}) = \prod_{j=1}^{n} P_{ji}$. Com o mesmo argumento anterior, temos que o estimador de máxima verossimilhança será $b_i = +\infty$. Problemas similares a esse ocorrem com os parâmetros a_i e c_i. No Capítulo 7, algumas formas de tratar esses problemas serão abordados.

3.2.5 Estimativas iniciais

Um ponto importante no processo de estimação é a obtenção de valores iniciais para os parâmetros. Se o item i tem m_i alternativas possíveis, um chute razoável para o parâmetro de acerto ao acaso é $c_i = 1/m_i$. Richardson (1936) e Tucker (1946) mostraram que se adotarmos a FRI Normal, então

$$\rho_{T,U_i} = \frac{a_i}{\sqrt{1 + a_i^2}}, \quad -1 < \rho_{T,U_i} < 1, \tag{3.30}$$

onde ρ_{T,U_i} é o coeficiente de correlação bisserial, utilizado na Teoria Clássica de Medidas. Este coeficiente é estimado pelo coeficiente de correlação de Pearson entre os escores, T_j, e as respostas ao item i. Com isso, obtemos $\widehat{a}_i^{(0)}$.

Em complemento, Tucker (1946) expressou o parâmetro de dificuldade associado ao item i da teoria clássica de itens π_i (proporção verdadeira de respostas corretas) como

$$\pi_i = \Phi(-\nu_i), \quad \nu_i = b_i \rho_{T,U_i}, \tag{3.31}$$

onde Φ é a função de distribuição associada à N(0,1). Vale notar que no caso de

usar a função Logística para a FRI, o fator $D = 1,702$ torna os modelos Normal e Logístico muito próximos (ver Halley (1952)) de forma que as expressões (3.30) e (3.31) produzem bons resultados para o modelo logístico.

3.3 Estimação das habilidades

Nesta seção vamos tratar da estimação das habilidades quando os parâmetros dos itens são conhecidos. Na prática, essa situação ocorre quando os itens já foram calibrados (estimados) em outros testes. Como a calibração dos itens deve ser feita com um número grande de indivíduos, a estimação das habilidades de um grupo pequeno de indivíduos é mais confiável se forem utilizados itens já calibrados.

Pela independência entre as respostas de diferentes indivíduos (S1) e a independência local (S2), podemos escrever a log-verossimilhança como em (3.3), agora como função de $\boldsymbol{\theta}$ e não de $\boldsymbol{\zeta}$, ou seja,

$$\log L(\boldsymbol{\theta}) \;=\; \sum_{j=1}^{n}\sum_{i=1}^{I}\{u_{ji}\log P_{ji} + (1-u_{ji})\log Q_{ji}\}. \tag{3.32}$$

O EMV de θ_j é o valor que maximiza a verossimilhança, ou equivalentemente, é a solução da equação

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j} = 0, \quad j = 1, \cdots, n. \tag{3.33}$$

Notemos, de (3.32), que

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j} &= \sum_{i=1}^{I} \left\{ u_{ji} \frac{\partial(\log P_{ji})}{\partial \theta_j} + (1-u_{ji}) \frac{\partial(\log Q_{ji})}{\partial \theta_j} \right\} && (3.34)\\
&= \sum_{i=1}^{I} \left\{ u_{ji} \frac{1}{P_{ji}} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) - (1-u_{ji}) \frac{1}{Q_{ji}} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) \right\} \\
&= \sum_{i=1}^{I} \left\{ u_{ji} \frac{1}{P_{ji}} - (1-u_{ji}) \frac{1}{Q_{ji}} \right\} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) \\
&= \sum_{i=1}^{I} \left\{ \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right\} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) && (3.35)\\
&= \sum_{i=1}^{I} \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right), && (3.36)
\end{aligned}$$

onde a última igualdade segue de (3.5). Como

$$\frac{\partial P_{ji}}{\partial \theta_j} = D a_i (1-c_i) P^*_{ji} Q^*_{ji}, \qquad (3.37)$$

obtêm-se

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j} &= \sum_{i=1}^{I} \left\{ (u_{ji} - P_{ji}) D a_i (1-c_i) P^*_{ji} Q^*_{ji} \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= D \sum_{i=1}^{I} a_i (1-c_i)(u_{ji} - P_{ji}) W_{ji}. && (3.38)
\end{aligned}$$

Segue então que a equação de estimação (3.33) para $\theta_j,\ j = 1, \cdots, n$, é

$$\theta_j \quad : \quad D \sum_{i=1}^{I} a_i (1-c_i)(u_{ji} - P_{ji}) W_{ji} = 0. \qquad (3.39)$$

Novamente, esta equação não apresenta solução explícita para θ_j e, por isso, precisamos de algum método iterativo para obter as estimativas desejadas. A seguir, obteremos as expressões necessárias para aplicações dos processos iterativos Newton-Raphson e "Scoring"de Fisher.

3.3.1 Aplicação do algoritmo Newton-Raphson

De forma similar ao que foi feito na Seção 3.2.1, e considerando $\widehat{\theta}_j^{(t)}$ a estimativa de θ_j na iteração t, então na iteração $t+1$ do algoritmo Newton-Raphson teremos que

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - [H(\widehat{\theta}_j^{(t)})]^{-1} h(\widehat{\theta}_j^{(t)}) \tag{3.40}$$

onde, ver Apêndice A.2 para as demonstrações dos resultados,

$$\begin{aligned} h(\theta_j) &\equiv \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j} \\ &= \sum_{i=1}^{I} \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} (P^*_{ji} Q^*_{ji}) h_{ji} \\ &= \sum_{i=1}^{I} (u_{ji} - P_{ji}) W_{ji} h_{ji} \end{aligned}$$

e

$$\begin{aligned} H(\theta_j) &\equiv \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j^2} \\ &= \sum_{i=1}^{I} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) (P^*_{ji} Q^*_{ji}) H_{ji} - \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right)^2 (P^*_{ji} Q^*_{ji})^2 h_{ji}^2 \right\} \\ &= \sum_{i=1}^{I} (u_{ji} - P_{ji}) W_{ji} \left\{ H_{ji} - (u_{ji} - P_{ji}) W_{ji} h_{ji}^2 \right\} \end{aligned} \tag{3.41}$$

com

$$h_{ji} = (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial P_{ji}}{\partial \theta_j}\right) = Da_i(1-c_i) \tag{3.42}$$

e

$$H_{ji} = (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2}\right) = D^2a_i^2(1-c_i)(1-2P^*_{ji}). \tag{3.43}$$

3.3.2 Aplicação do método "Scoring"de Fisher

Para aplicação do método "Scoring"de Fisher, devemos substituir os componentes da matriz de derivadas segundas usadas no processo iterativo de Newton-Raphson pelos seus valores esperados. Por (3.41) temos que

$$\begin{aligned}\Delta(\theta_j) &\equiv E(H(\theta_j)) \\ &= \sum_{i=1}^{I}\left\{E(U_{ji}-P_{ji})W_{ji}H_{ji} - E(U_{ji}-P_{ji})^2W_{ji}^2h_{ji}^2\right\} \\ &= -\sum_{i=1}^{I}P^*_{ji}Q^*_{ji}W_{ji}h_{ji}^2. \end{aligned} \tag{3.44}$$

Neste caso, a expressão para estimativa de $\theta_j, j = 1, \cdots, n$, na iteração $t+1$ será

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - [\Delta(\widehat{\theta}_j^{(t)})]^{-1}h(\widehat{\theta}_j^{(t)}).$$

3.3.3 Erro-padrão

Sob algumas condições de regularidade (ver Sen & Singer (1993), por exemplo) a distribuição assintótica do estimador de máxima verossimilhança, $\widehat{\theta}_j$, é normal com média θ_j e variância dada pela inversa da matriz de informação

$$I(\theta_j) = -E\left(\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j^2}\right) = -\Delta(\theta_j), \tag{3.45}$$

onde $\Delta(\theta_j)$ é obtida de (3.44). A raiz quadrada de $I(\theta_j)$ fornece o erro-padrão de $\widehat{\theta}_j$.

3.3.4 Escore nulo ou perfeito

Tal como na estimação dos parâmetros dos itens, existe um problema a ser contornado na estimação por máxima verossimilhança. Se o indivíduo j obtém escore nulo, isto é, $u_{ji} = 0,\ i = 1, \cdots, I$, então a verossimilhança resume-se a $L(\boldsymbol{\theta}) = \prod_{i=1}^{I} Q_{ji}$. Como $Q_{ji},\ i = 1, \cdots, I$, é decrescente com θ_j, então $L(\boldsymbol{\theta})$ tambem é decrescente com θ_j e daí o estimador de máxima verossimilhança será $\theta_j = -\infty$. Por outro lado, se o indivíduo j obter o escore total, isto é, $u_{ji} = 1,\ i = 1, \cdots, I$, então a verossimilhança resume-se a $L(\boldsymbol{\theta}) = \prod_{i=1}^{I} P_{ji}$. Como $P_{ji},\ i = 1, \cdots, I$, é crescente com θ_j, então $L(\boldsymbol{\theta})$ tambem é crescente com θ_j e daí o estimador de máxima verossimilhança será $\theta_j = +\infty$. Algumas formas de como tratar esse problema serão abordadas ainda nesse capítulo e no Capítulo 7.

3.3.5 Estimativas iniciais

A obtenção de estimativas (valores) iniciais para o início do processo de estimação pode ser feita com os escores padronizados. Se T_j é o escore do indivíduo j, m o escore médio e s o desvio-padrão dos escores dos n indivíduos, então $\widehat{\theta}_j^{(0)} = (T_j - m)/s$.

3.4 Estimação conjunta: parâmetros dos itens e habilidades

Nesta etapa trataremos do caso mais comum, em que nem os parâmetros dos itens e nem as habilidades são conhecidos. As Seções 3.2 e 3.3 compõem as partes básicas da estimação conjunta. Nas expressões (3.14) a (3.16) e (3.39) temos as equações de estimação para os parâmetros dos itens e habilidades. Entretanto, estas equações não apresentam expressões explicitas para os respectivos EMV. Por conta disso, algum processo iterativo deve ser aplicado no processo de maximização e, como conseqüência, algumas quantidades ou suposições podem ser adicionadas ao modelo.

A principal diferença da estimação conjunta se dá no tratamento da métrica (escala) em que todos os parâmetros são estimados. Quando tratamos da estimação dos parâmetros dos itens com as habilidades conhecidas, não houve necessidade do arbítrio da métrica, pois estes são estimados na métrica das habilidades. Por outro lado, quando tratamos da estimação das habilidades com os parâmetros dos itens conhecidos, estas são estimadas na métrica dos parâmetros dos itens. Na estimação conjunta não há uma métrica definida e, portanto, deveremos estabelecê-la. A explicação formal para a necessidade do estabelecimento da métrica dos parâmetros consiste em um problema denominado *falta de identificabilidade do modelo*. Essa não-identificabilidade ocorre porque mais de um conjunto de parâmetros produz o mesmo valor no ML3, e consequentemente, na verossimilhança. Conforme já citado no Capítulo 2, se $\theta_j^* = \alpha\theta_j + \beta$, $b_i^* = \alpha b_i + \beta$, $a_i^* = a_i/\alpha$ e $c_i^* = c_i$, onde α e β são constantes reais com $\alpha > 0$, então

$$\begin{aligned} P(U_{ji} = 1|\theta_j^*, \boldsymbol{\zeta}_i^*) &= c_i^* + (1 - c_i^*)\{1 + \exp[-Da_i^*(\theta_j^* - b_i^*)]\}^{-1} \\ &= c_i + (1 - c_i)\{1 + \exp\left[-D\frac{a_i}{\alpha}(\alpha\theta_j + \beta - (\alpha b_i + \beta))\right]\}^{-1} \\ &= c_i + (1 - c_i)\{1 + \exp[-Da_i(\theta_j - b_i)]\}^{-1} \\ &= P(U_{ji} = 1|\theta_j, \boldsymbol{\zeta}_i). \end{aligned}$$

Essa não-identificabilidade pode ser eliminada de várias formas, como fixando alguns valores para as habilidades, por exemplo. Entretanto, devemos ressaltar que essa não-identificabilidade está intimamente relacionada à características da população envolvida no estudo. Até agora não especificamos quando uma habilidade pode ser considerada alta ou baixa, nem como diagnosticar o quanto uma habilidade está afastada de outra. Isso pode ser resolvido especificando uma medida de posição (média, por exemplo) e outra de dispersão (desvio-padrão, por exemplo) para as habilidades. Dessa forma estaremos definindo uma *métrica* (unidade de medida) para as habilidades e, consequentemente, para os parâmetros dos itens. De forma geral, podemos dizer que estamos trabalhando com variáveis latentes, e nessa situação sempre há a necessidade do estabelecimento da métrica. Neste livro, vamos eliminar o problema de não-identificabilidade do modelo padronizando as habilidades de forma que estas tenham uma média especificada μ e desvio-padrão σ. Desta

forma, as habilidades e os parâmetros dos itens são estimados na métrica (μ, σ). Em muitas situações adota-se $\mu = 0$ e $\sigma = 1$, valores que serão considerados em todo o restante do livro.

Para aplicação do algoritmo Newton-Raphson são necessárias as derivadas segundas da log-verossimilhança, com relação a $\boldsymbol{\zeta}_i$ e θ_j, $i = 1, \cdots, I$ e $j = 1, \cdots, n$. Estas derivadas compõem uma matriz $\boldsymbol{H}$ quadrada de ordem $(3I + n)$ e essa dimensão pode ser suficientemente grande de forma a causar uma enorme exigência computacional. Por isso, precisamos explorar um pouco mais a estrutura de $\boldsymbol{H}$. Notemos que pela independência local, temos

$$\frac{\partial L(\boldsymbol{\zeta}, \boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_l'} = \mathbf{0}, \quad \text{para } i \neq l. \tag{3.46}$$

Pela independência entre as respostas de indivíduos diferentes, temos que

$$\frac{\partial L(\boldsymbol{\zeta}, \boldsymbol{\theta})}{\partial \theta_j \partial \theta_l} = 0, \quad \text{para } j \neq l. \tag{3.47}$$

Vale notar que (3.46) e (3.47) são conseqüências das suposições inerentes do modelo. Uma suposição adicional que simplifica bastante a estrutura de $\boldsymbol{H}$ é a de que não existe correlação entre itens e habilidades. Essa suposição condiz com situações práticas, pois as habilidades são inerentes dos indivíduos, que em nada dependem dos itens envolvidos no estudo. Como conseqüência desta suposição, temos que

$$\frac{\partial L(\boldsymbol{\zeta}, \boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_i \partial \theta_j} = \mathbf{0}, \quad \text{para } i = 1, \cdots, I \text{ e } j = 1, \cdots, n. \tag{3.48}$$

Assim, a matriz $\boldsymbol{H}$ torna-se bloco-diagonal, na qual os I primeiros blocos são matrizes 3×3 relativas aos parâmetros dos itens e os n blocos seguintes são escalares relativos às habilidades. As expressões (3.46) a (3.48) facilitam bastante a estrutura de $\boldsymbol{H}$, mas não diminuem sua dimensão. Entretanto, com base nessa estrutura bloco-diagonal, Birbaum (1968) propôs um algoritmo em que os itens e as habilidades são estimados individualmente, utilizando o algoritmo Newton-Raphson ou o método "Scoring"de Fisher, no qual cada iteração é composta de dois estágios:

Estágio 1: Começando com estimativas iniciais para as habilidades $\boldsymbol{\theta}$ (escores padronizados, por exemplo) e tratando estas habilidade como conhecidas, estimamos $\boldsymbol{\zeta}_i,\ i = 1, \cdots, I$.

Estágio 2: Começando com estimativas iniciais (obtidas no Estágio 1) para $\boldsymbol{\zeta}$ e tratando estes parâmetros como conhecidos, estimamos as habilidades $\theta_j,\ j = 1, \cdots, n$.

No Estágio 1, os itens são estimados empregando o desenvolvimento da Seção 3.2. No Estagio 2 as habilidades são estimadas com a teoria desenvolvida na Seção 3.3. Este processo de dois estágios é repetido até a convergência das habilidades e dos parâmetros dos itens.

Comentários

Os erros-padrão para $\widehat{\boldsymbol{\zeta}}_i,\ i = 1, \cdots, I$, e $\widehat{\theta}_j,\ j = 1, \cdots, n$, continuam sendo obtidos com o uso das expressões (3.29) e (3.45). Além disso, a estimação conjunta apresenta os mesmos problemas já citados anteriormente, ou seja, quando algum item é respondido corretamente, ou incorretamente, por todos os indivíduos, ou quando algum indivíduo responde corretamente, ou incorretamente, a todos os itens. Mais adiante, nesse capítulo e no Capítulo 7, veremos como tratar destes casos.

3.5 Máxima verossimilhança marginal

O método da Máxima Verossimilhança Marginal, proposto por Bock & Lieberman (1970 apresenta algumas vantagens em relação ao método da Máxima Verossimilhança Conjunta. A proposta desse método é fazer a estimação em duas etapas: primeiro os parâmetros dos itens e, posteriormente, as habilidades. Como as habilidades não são conhecidas, precisaremos usar algum artifício de forma que a verossimilhança não seja mais função das habilidades. Anderson (1980) argumenta que se considerarmos uma população Π composta por n indivíduos com habilidades $\theta_j,\ j = 1, \cdots, n$, e construirmos a distribuição de frequência acumulada $G(\theta)$=(número de $j : \theta_j \leq \theta)/n$, então, se n for suficientemente grande os θ_j estarão bastante próximos de forma que $G(\theta)$ pode ser aproximada por uma distribuição contínua. A densidade $g(\theta)$, relativa à $G(\theta)$,

pode realmente ser considerada a função densidade para θ no experimento de retirar um indivíduo ao acaso da população Π e observar seu parâmetro θ. Neste contexto, é importante ressaltar que, quando atribuímos uma distribuição de probabilidade para θ *não estamos aplicando nenhum argumento bayesiano*. A distribuição de θ realmente existe, no sentido explicado acima, como a densidade relativa à distribuição $G(\theta)$.

De acordo com isso, um artifício para eliminar as habilidades na verossimilhança consiste em marginalizar a verossimilhança integrando-a com relação à distribuição da habilidade. De forma geral, consideremos que as habilidades, $\theta_j,\ j = 1, \cdots, n$, são realizações de uma variável aleatória θ com distribuição contínua e função densidade de probabilidade (fdp) $g(\theta|\boldsymbol{\eta})$, duplamente diferenciável, com as componentes de $\boldsymbol{\eta}$ conhecidas e finitas. Para o caso em que θ tem distribuição Normal, temos $\boldsymbol{\eta} = (\mu, \sigma^2)$, onde μ é a média e σ^2 a variância das habilidades dos indivíduos de Π. Portanto, se desejarmos que os itens sejam estimados na métrica (0,1), deveremos adotar $\mu = 0$ e $\sigma = 1$.

3.5.1 Abordagem de Bock & Lieberman

Com as notações acima, temos que a probabilidade marginal de $\boldsymbol{U}_{j.}$ é dada por

$$\begin{aligned} P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta}) &= \int_{I\!R} P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta}, \boldsymbol{\eta}) g(\theta|\boldsymbol{\eta}) d\theta \\ &= \int_{I\!R} P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta}) g(\theta|\boldsymbol{\eta}) d\theta, \end{aligned} \tag{3.49}$$

onde na última igualdade usamos que a distribuição de $\boldsymbol{U}_{j.}$ não é função de $\boldsymbol{\eta}$ e $I\!R$ representa o conjunto dos números reais. Usando a independência entre as respostas de diferentes indivíduos, podemos escrever a probabilidade associada ao vetor de respostas $\boldsymbol{U}_{..}$ como

$$P(\boldsymbol{u}_{..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) = \prod_{j=1}^{n} P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta}). \tag{3.50}$$

Embora a verossimilhança possa ser escrita como (3.50), tem sido freqüente utilizar a abordagem de *Padrões de Resposta*. Como temos I itens no total, com 2 possíveis respostas para cada item (0 ou 1), há $S = 2^I$ possíveis respostas

(padrões de resposta). Quando o número de indivíduos é grande com relação ao número de itens, pode haver vantagens computacionais em trabalhar com o número de ocorrências dos diferentes padrões de resposta. Neste sentido, daqui em diante vamos trabalhar considerando este raciocínio. O índice j não mais representará um indivíduo, mas sim um padrão de resposta.

Seja r_j o número de ocorrências distintas do padrão de resposta j, e ainda $s \leq \min(n, S)$ o número de padrões de resposta com $r_j > 0$. Segue disso que

$$\sum_{j=1}^{s} r_j = n. \tag{3.51}$$

Pela independência entre as respostas dos diferentes indivíduos, temos que os dados seguem uma distribuição *Multinomial*, isto é,

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \frac{n!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} \left[P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta})\right]^{r_j}, \tag{3.52}$$

e, portanto, a log-verossimilhança é

$$\log L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \log \left\{ \frac{n!}{\prod_{j=1}^{s} r_j!} \right\} + \sum_{j=1}^{s} r_j \log P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta}).$$

As equações de estimação para os parâmetros dos itens são dadas por

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \mathbf{0}, \quad i = 1, \cdots, I, \tag{3.53}$$

com

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} &= \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left\{ \sum_{j=1}^{s} r_j \log P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta}) \right\} \\ &= \sum_{j=1}^{s} r_j \frac{1}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i}. \end{aligned} \tag{3.54}$$

Mas

$$\begin{aligned}\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} &= \frac{\partial}{\partial \boldsymbol{\zeta}_i}\int_{I\!R} P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta \\ &= \int_{I\!R}\left(\frac{\partial}{\partial \boldsymbol{\zeta}_i}P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})\right)g(\theta|\boldsymbol{\eta})d\theta \qquad (3.55)\\ &= \int_{I\!R}\left(\frac{\partial}{\partial \boldsymbol{\zeta}_i}\prod_{l=1}^{I}P(u_{jl}|\theta,\boldsymbol{\zeta}_l)\right)g(\theta|\boldsymbol{\eta})d\theta\end{aligned}$$

$$\begin{aligned}\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} &= \int_{I\!R}\left(\prod_{l\neq i}^{I}P(u_{jl}|\theta,\boldsymbol{\zeta}_l)\right)\left(\frac{\partial}{\partial \boldsymbol{\zeta}_i}P(u_{ji}|\theta,\boldsymbol{\zeta}_i)\right)g(\theta|\boldsymbol{\eta})d\theta \\ &= \int_{I\!R}\left(\frac{\partial P(u_{ji}|\theta,\boldsymbol{\zeta}_i)/\partial \boldsymbol{\zeta}_i}{P(u_{ji}|\boldsymbol{\zeta}_i)}\right)P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta, \qquad (3.56)\end{aligned}$$

onde a ordem da derivada e da integral em (3.55) pôde ser permutada com base no Teorema da Convergência Dominada de Lebesgue (Chow & Teicher, 1978). Reescrevendo $P(u_{ji}|\theta,\boldsymbol{\zeta}_i)$ como em (3.1), teremos que

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\zeta}_i}P(u_{ji}|\theta,\boldsymbol{\zeta}_i) &= \frac{\partial}{\partial \boldsymbol{\zeta}_i}\left(P_i^{u_{ji}}Q_i^{1-u_{ji}}\right) \\ &= u_{ji}P_i^{u_{ji}-1}\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)Q_i^{1-u_{ji}} + (1-u_{ji})Q_i^{-u_{ji}}\left(-\frac{\partial}{\partial \boldsymbol{\zeta}_i}P_i\right)P_i^{u_{ji}} \\ &= \left(u_{ji}P_i^{u_{ji}-1}Q_i^{1-u_{ji}} - (1-u_{ji})Q_i^{-u_{ji}}P_i^{u_{ji}}\right)\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}.\end{aligned}$$

Notemos agora que o termo entre parênteses vale 1 quando $u_{ji}=1$ e vale -1 quando $u_{ji}=0$, portanto podemos reescrevê-lo como $(-1)^{u_{ji}+1}$. Com isso,

$$\frac{\partial}{\partial \boldsymbol{\zeta}_i}P(u_{ji}|\theta,\boldsymbol{\zeta}_i) = (-1)^{u_{ji}+1}\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right). \qquad (3.57)$$

Notando agora que

$$\frac{(-1)^{u_{ji}+1}P_iQ_i}{P_i^{u_{ji}}Q_i^{1-u_{ji}}} = \begin{cases} Q_i & \text{se } u_{ji}=1 \\ -P_i & \text{se } u_{ji}=0, \end{cases} \tag{3.58}$$

podemos reescrever este termo como $u_{ji}-P_i$. Segue que (3.56) pode ser escrita como

$$\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \int_{I\!R}\left[\frac{(u_{ji}-P_i)}{P_iQ_i}\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)\right]P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta \tag{3.59}$$

Por conveniência, consideremos a seguinte ponderação:

$$W_i = \frac{P_i^*Q_i^*}{P_iQ_i}, \tag{3.60}$$

onde

$$P_i^* = \{1+e^{-Da_i(\theta-b_i)}\}^{-1} \quad \text{e} \quad Q_i^* = 1-P_i^*. \tag{3.61}$$

Com isso, podemos reescrever a Equação (3.59) como

$$\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \int_{I\!R}\left[(u_{ji}-P_i)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)\frac{W_i}{P_i^*Q_i^*}\right]P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta. \tag{3.62}$$

Usando a notação

$$g_j^*(\theta) \equiv g(\theta|\boldsymbol{u}_{j.},\boldsymbol{\zeta},\boldsymbol{\eta}) = \frac{P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}, \tag{3.63}$$

teremos que a função de verossimilhança (3.54) pode ser escrita como

$$\frac{\partial \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \sum_{j=1}^{s} r_j \int_{I\!R}\left[(u_{ji}-P_i)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)\frac{W_i}{P_i^*Q_i^*}\right]g_j^*(\theta)d\theta. \tag{3.64}$$

Resta agora a obtenção das equações específicas para cada parâmetro do vetor $\boldsymbol{\zeta}_i = (a_i, b_i, c_i)'$. As expressões para as derivadas de P_i são dadas por (3.8) a (3.10) com P_{ji}, Q_{ji}, P^*_{ji} e Q^*_{ji} substituídas por P_i, Q_i, P^*_i e Q^*_i, respectivamente.

Para obtermos a equação de estimação para o parâmetro de discriminação, a_i, notemos que da expressão (3.64) temos que

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial a_i} = & \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial a_i} \right) \frac{W_i}{P^*_i Q^*_i} \right] g^*_j(\theta) d\theta \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) D(1 - c_i)(\theta - b_i) P^*_i Q^*_i \frac{W_i}{P^*_i Q^*_i} \right] g^*_j(\theta) d\theta \\
&= D(1 - c_i) \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i)(\theta - b_i) W_i \right] g^*_j(\theta) d\theta. \qquad (3.65)
\end{aligned}$$

Para o parâmetro de dificuldade, b_i, temos que

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial b_i} = & \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial b_i} \right) \frac{W_i}{P^*_i Q^*_i} \right] g^*_j(\theta) d\theta \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i)(-1) D a_i (1 - c_i) P^*_i Q^*_i \frac{W_i}{P^*_i Q^*_i} \right] g^*_j(\theta) d\theta \\
&= -D a_i (1 - c_i) \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) W_i \right] g^*_j(\theta) d\theta. \qquad (3.66)
\end{aligned}$$

Para o parâmetro de acerto ao acaso, c_i, temos que

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial c_i} &= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)\left(\frac{\partial P_i}{\partial c_i}\right)\frac{W_i}{P_i^* Q_i^*}\right] g_j^*(\theta)d\theta \\ &= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)Q_i^*\frac{W_i}{P_i^* Q_i^*}\right] g_j^*(\theta)d\theta \\ &= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)\frac{W_i}{P_i^*}\right] g_j^*(\theta)d\theta. \qquad (3.67)\end{aligned}$$

Em resumo, as equações de estimação para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i \quad : \quad D(1-c_i)\sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)(\theta-b_i)W_i\right] g_j^*(\theta)d\theta = 0, \quad (3.68)$$

$$b_i \quad : \quad -Da_i(1-c_i)\sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)W_i\right] g_j^*(\theta)d\theta = 0, \quad (3.69)$$

$$c_i \quad : \quad \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)\frac{W_i}{P_i^*}\right] g_j^*(\theta)d\theta = 0, \quad (3.70)$$

as quais não possuem solução explícita.

3.5.2 Métodos iterativos

Para aplicação do algoritmo Newton-Raphson, precisaremos das derivadas segundas de $\log L(\boldsymbol{\zeta},\boldsymbol{\eta})$. Quando desenvolvemos as expressões para a estimação de $\boldsymbol{\zeta}_i$ na Seção 3.2, a propriedade de independência local foi suficiente para garantir que os (parâmetros dos) itens pudessem ser estimados individualmente, pois a derivada segunda de $\log L(\boldsymbol{\zeta})$ com relação a $\boldsymbol{\zeta}_i$ e $\boldsymbol{\zeta}_l$, para $l \neq i$, era nula. Entretanto, na estimação por máxima verossimilhança marginal isso não acontece, levando à necessidade da estimação dos I itens conjuntamente. As expressões para as derivadas segundas são obtidas a partir de

$$\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}_l\partial\boldsymbol{\zeta}_i'} &= \frac{\partial}{\partial\boldsymbol{\zeta}_l}\left[\frac{\partial \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}_i}\right]' \\
&= \frac{\partial}{\partial\boldsymbol{\zeta}_l}\left[\sum_{j=1}^{s} r_j \frac{1}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}_i}\right]' \\
&= \sum_{j=1}^{s} r_j \left\{\frac{\partial^2 P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})/(\partial\boldsymbol{\zeta}_l\partial\boldsymbol{\zeta}_i')}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})} - \left(\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})/\partial\boldsymbol{\zeta}_l}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}\right)\left(\frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})/\partial\boldsymbol{\zeta}_i}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}\right)'\right\}.
\end{aligned} \tag{3.71}$$

para $i, l = 1, \cdots, I$.

Considerando $\widehat{\boldsymbol{\zeta}}^{(t)}$ a estimativa de $\boldsymbol{\zeta}$ na iteração t, então na iteração $t+1$ teremos que

$$\widehat{\boldsymbol{\zeta}}^{(t+1)} = \widehat{\boldsymbol{\zeta}}^{(t)} - [\boldsymbol{H}_{PI}(\widehat{\boldsymbol{\zeta}}^{(t)})]^{-1}\boldsymbol{h}_{PI}(\widehat{\boldsymbol{\zeta}}^{(t)}) \tag{3.72}$$

onde

$$\boldsymbol{h}_{PI}(\boldsymbol{\zeta}) = \begin{pmatrix} \boldsymbol{h}(\boldsymbol{\zeta}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{\zeta}_I) \end{pmatrix} \qquad \text{e} \qquad \boldsymbol{H}_{PI}(\boldsymbol{\zeta}) = \begin{pmatrix} \boldsymbol{H}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_1) & \cdots & \boldsymbol{H}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_I) \\ \vdots & \vdots & \vdots \\ \boldsymbol{H}(\boldsymbol{\zeta}_I,\boldsymbol{\zeta}_1) & \cdots & \boldsymbol{H}(\boldsymbol{\zeta}_I,\boldsymbol{\zeta}_I) \end{pmatrix},$$

com

$$\begin{aligned}
\boldsymbol{h}(\boldsymbol{\zeta}_i) &= \frac{\partial \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}_i} \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} (u_{ji} - P_i) W_i \boldsymbol{h}_i g_j^*(\boldsymbol{\theta}) d\theta,
\end{aligned} \tag{3.73}$$

e

$$\begin{aligned}
\boldsymbol{H}(\boldsymbol{\zeta}_i,\boldsymbol{\zeta}_l) &= \frac{\partial^2 \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}_l\partial\boldsymbol{\zeta}_i'} \\
&= \sum_{j=1}^{s} r_j \left\{\boldsymbol{H}_{il(j)} - \boldsymbol{h}_{i(j)}\boldsymbol{h}_{l(j)}'\right\}.
\end{aligned} \tag{3.74}$$

No Apêndice A.3 o leitor encontrará as demonstrações para os resultados acima.

Para aplicarmos o algoritmo "Scoring"de Fisher, notemos que $E[H_{il(j)}] = 0,\ i,l = 1,\cdots,I$ e $j = 1,\cdots,n$. Segue então que

$$\boldsymbol{\Delta}(\boldsymbol{\zeta}_i,\boldsymbol{\zeta}_l) = E[H(\boldsymbol{\zeta}_i,\boldsymbol{\zeta}_l)] = -\sum_{j=1}^{s} r_j[h_{i(j)}h_{i(j)}].$$

3.5.3 Métodos de quadratura

Antes de prosseguir, vamos discutir um problema importante encontrado na implementação da estimação dos parâmetros dos itens. Podemos notar que as equações (3.68) a (3.70), por exemplo, envolvem integrais que não apresentam solução analítica. Por conta disso, algum meio deve ser encontrado para a solução (aproximação) numérica de uma integral. Embora existam muitos métodos de aproximações de integrais, na TRI têm sido freqüente a aplicação do método *Hermite-Gauss*, usualmente denominado de *método de quadratura gaussiana*. Se $g(\theta|\boldsymbol{\eta})$ é uma função contínua com integral finita, ela pode ser aproximada, para qualquer grau de precisão, por uma outra função que assume um número finito de pontos. Dessa forma, o problema de obter a integral de uma função contínua é substituído pela obtenção da soma das área de um número finito, digamos q, de retângulos. Os pontos médios de cada retângulo, $\overline{\theta}_k,\ k = 1,\cdots,q$, são denominados de *nós* (ou *pontos de quadratura*). Cada nó tem um peso $A_k \equiv A(\overline{\theta}_k)$ associado que leva em conta a altura $g(\overline{\theta}_k|\boldsymbol{\eta})$ e a largura (Δ_k) do respectivo intervalo, tal como $A_k = g(\overline{\theta}_k|\boldsymbol{\eta}) \times \Delta_k$. Os valores $\overline{\theta}_k$ e A_k são obtidos resolvendo-se um conjunto de equações que envolvem a função $g(\theta|\boldsymbol{\eta})$ e o número de nós (ver Hildebrand (1956), páginas 327-330). Uma tabela para $\overline{\theta}_k$ e A_k relativa a função gaussiana pode ser encontrada em Stroud & Sechest (1966). Para adaptar essa tabela para o caso em que $g(\theta|\boldsymbol{\eta})$ representa a fdp de uma variável $N(0,1)$, basta multiplicar os nós $\overline{\theta}_k$ por $\sqrt{2}$ e dividir os pesos A_k por $\sqrt{\pi}$.

Equações de estimação em forma de quadratura

Consideremos conhecidos os nós $\overline{\theta}_k$ e os pesos, A_k, $k = 1, \cdots, q$, com $A_k = g(\overline{\theta}_k|\boldsymbol{\eta}) \times \Delta_k$. Com isso, podemos escrever

$$P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) = \prod_{i=1}^{I} [P_{ki}^{u_{ji}} Q_{ki}^{1-u_{ji}}],$$

$$P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) g(\overline{\theta}_k|\boldsymbol{\eta}) = P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) A_k \Delta_k^{-1},$$

$$P(\boldsymbol{u}_{j\cdot}|\boldsymbol{\zeta}, \boldsymbol{\eta}) \simeq \sum_{k=1}^{q} P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) g(\overline{\theta}_k|\boldsymbol{\eta}) \Delta_k = \sum_{k=1}^{q} P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) A_k.$$

Segue que (3.63) pode ser escrita, em forma de quadratura, como

$$g_j^*(\overline{\theta}_k) \simeq \frac{P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) A_k}{\sum_{k=1}^{q} P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) A_k} \Delta_k^{-1}. \tag{3.75}$$

Por exemplo, voltando à função de verossimilhança para a_i dada por (3.68), podemos reescrevê-la em forma de quadratura como

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial a_i} &= D(1 - c_i) \sum_{j=1}^{s} r_j \int_{\mathbb{R}} [(u_{ji} - P_i)(\theta - b_i) W_i] \, g_j^*(\theta) d\theta \\ &\simeq D(1 - c_i) \sum_{j=1}^{s} \sum_{k=1}^{q} r_j \left[(u_{ji} - P_{ki})(\overline{\theta}_k - b_i) W_{ki} \right] g_j^*(\overline{\theta}_k) \Delta_k \end{aligned}$$

Para que a expressão em forma de quadratura fique o mais parecida possível com a original, podemos redefinir a quantidade $g_j^*(\theta_k)$ de (3.75) por

$$g_j^*(\theta_k) = \frac{P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) A_k}{\sum_{k=1}^{q} P(\boldsymbol{u}_{j\cdot}|\overline{\theta}_k, \boldsymbol{\zeta}) A_k}. \tag{3.76}$$

Desta forma, a função de verossimilhança para a_i fica

$$a_i \quad : \quad D(1-c_i)\sum_{j=1}^{s}\sum_{k=1}^{q} r_j \left[(u_{ji}-P_{ki})(\overline{\theta}_k - b_i)W_{ki}\right] g_j^*(\overline{\theta}_k). \qquad (3.77)$$

De forma análoga, temos que as equações de estimação em forma de quadratura para os parâmetros b_i e c_i são, respectivamente,

$$b_i \quad : \quad -Da_i(1-c_i)\sum_{j=1}^{s}\sum_{k=1}^{q} r_j \left[(u_{ji}-P_{ki})W_{ki}\right] g_j^*(\overline{\theta}_k) = 0, \qquad (3.78)$$

$$c_i \quad : \quad \sum_{j=1}^{s}\sum_{k=1}^{q} r_j \left[(u_{ji}-P_{ki})\frac{W_{ki}}{P_{ki}^*}\right] g_j^*(\overline{\theta}_k) = 0. \qquad (3.79)$$

Deve ser ressaltado que a função $g_j^*(\overline{\theta}_k)$ nas equações (3.77) a (3.79) deve ser calculada por (3.76). Novamente, estas equações não apresentam soluções explícitas para os EMV dos parâmetros dos itens. Para aplicação dos procedimentos Newton-Raphson ou "Scoring"de Fisher devemos notar que as derivadas segundas de $\log L(\boldsymbol{\zeta}, \boldsymbol{\eta})$ com relação a $\boldsymbol{\zeta}_i$ e $\boldsymbol{\zeta}_l$, para $i \neq l$, não são nulas, o que leva à necessidade da estimação dos parâmetros dos I itens simultaneamente. Isso pode gerar uma grande limitação na estimação de um número alto de itens devido à necessidade da inversão de matrizes de dimensões $3I \times 3I$. A proposta de Bock & Aitkin (1981), que apresentaremos a seguir, contorna este problema.

3.5.4 Abordagem de Bock & Aitkin

Uma reformulação da abordagem de Bock & Lieberman, que foi considerada satisfatória do ponto de vista computacional, foi proposta por Bock & Aitkin (1981). Esta reformulação teve como base a suposição de que os itens são independentes, de forma que

$$\frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_l} = \mathbf{0}, \quad \text{para } i \neq l. \qquad (3.80)$$

Essa suposição modifica a matriz $\boldsymbol{H}_{PI}(\boldsymbol{\zeta})$ tornando-a bloco diagonal, uma

situação similar à da Seção 3.4 onde eram estimados os parâmetros dos itens e as habilidades conjuntamente. Naquele caso, a independência local foi suficiente para garantir (3.80) e, assim, possibilitar que os itens fossem estimados individualmente, fixadas as habilidades. A proposta de Bock & Aitkin foi adotar a independência entre os itens de forma a possibilitar que os itens sejam estimados individualmente. Vale notar que as suposições de independência local e a suposição de independência dos itens são completamente diferentes. A primeira está relacionada às respostas dos indivíduos, enquanto a segunda se refere apenas aos itens.

Com esta construção, a estimação pode ser feita adotando as mesmas expressões desenvolvidas na seção anterior, fazendo a adaptação devida a (3.80). Entretanto, Bock & Aitkin sugerem que a obtenção das estimativas de máxima verossimilhança seja feita através do algorítimo EM, introduzido por Dempster, Laird & Rubin (1977), e por isso algumas alterações nas expressões da seção anterior serão necessárias.

De (3.68) temos que

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial a_i} &= D(1-c_i)\sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji}-P_i)(\theta-b_i)W_i\right] g_j^*(\theta)d\theta \\ &= D(1-c_i)\sum_{j=1}^{s} r_j \int_{I\!R} (\theta-b_i)\left[u_{ji}g_j^*(\theta)-P_i g_j^*(\theta)\right] W_i d\theta\end{aligned}$$

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial a_i} &= D(1-c_i)\int_{I\!R} (\theta-b_i)\left[\sum_{j=1}^{s} r_j u_{ji} g_j^*(\theta) - P_i \sum_{j=1}^{s} r_j g_j^*(\theta)\right] W_i d\theta \\ &= D(1-c_i)\int_{I\!R} (\theta-b_i)\left[r_i(\theta)-P_i f_i(\theta)\right] W_i d\theta, \qquad (3.81)\end{aligned}$$

onde

$$r_i(\theta)=\sum_{j=1}^{s} r_j u_{ji} g_j^*(\theta), \quad f_i(\theta)=\sum_{j=1}^{s} r_j g_j^*(\theta).$$

Lembrando que $g_j^*(\theta)$ é a distribuição condicional de θ_j dado $\boldsymbol{u}_{j.}$, então $f_i(\theta)$ representa o número esperado de indivíduos, dentre os que responderam o item i em uma população de tamanho n, que têm habilidade θ. Para a quantidade $r_i(\theta)$ contribuem apenas os indivíduos que responderam corretamente ao item i. Logo, esta quantidade representa o número esperado de indivíduos, dentre os que responderam corretamente ao item i em uma população de tamanho n, que têm habilidade θ.

Analogamente, de (3.69) e (3.70) temos que

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial b_i} = -Da_i(1-c_i)\int_{I\!R}\left[r_i(\theta) - P_i f_i(\theta)\right] W_i d\theta, \tag{3.82}$$

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial c_i} = \int_{I\!R}\left[r_i(\theta) - P_i f_i(\theta)\right] W_i d\theta. \tag{3.83}$$

Equações de estimação em forma de quadratura

Considerando conhecidos os nós $\overline{\theta}_k$ e os pesos, A_k, $k = 1, \cdots, q$, temos que as equações de estimação em forma de quadratura para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i \;:\; D(1-c_i)\sum_{k=1}^{q}(\overline{\theta}_k - b_i)\left[r_{ki} - P_{ki}f_{ki}\right]W_{ki} = 0, \tag{3.84}$$

$$b_i \;:\; -Da_i(1-c_i)\sum_{k=1}^{q}\left[r_{ki} - P_{ki}f_{ki}\right]W_{ki} = 0, \tag{3.85}$$

$$c_i \;:\; \sum_{k=1}^{q}\left[r_{ki} - P_{ki}f_{ki}\right]\frac{W_{ki}}{P^*_{ki}} = 0. \tag{3.86}$$

onde

$$r_{ki} = \sum_{j=1}^{s} r_j u_{ji} g^*_{jk}, \quad f_{ki} = \sum_{j=1}^{s} r_j g^*_{jk} \quad \text{e} \quad g^*_{jk} = g^*_j(\overline{\theta}_k). \tag{3.87}$$

3.5.5 Aplicação do algoritmo EM

O algoritmo EM é um processo iterativo para determinação de estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade na presença de variáveis aleatórias não observadas. Cada iteração deste processo é feita em dois passos: Esperança (E) e Maximização (M). No caso da TRI, o objetivo é obter estimativas de $\boldsymbol{\zeta}$ na presença das variáveis não observadas $\boldsymbol{\theta}$. Neste caso, $\boldsymbol{u}_{\cdot\cdot}$ representa o vetor de dados incompletos e $(\boldsymbol{u}_{\cdot\cdot}, \boldsymbol{\theta})$ o vetor de dados completos. Seja $f(\boldsymbol{u}_{\cdot\cdot}, \boldsymbol{\theta}|\boldsymbol{\zeta})$ a densidade conjunta do dados completos. Se $\widehat{\boldsymbol{\zeta}}^{(k)}$ é uma estimativa de $\boldsymbol{\zeta}$ na iteração t, então os passos EM para obtenção de $\widehat{\boldsymbol{\zeta}}^{(k+1)}$ são

Passo E: Calcular $E[\log f(\boldsymbol{u}_{\cdot\cdot}, \boldsymbol{\theta}|\boldsymbol{\zeta})|\boldsymbol{u}_{\cdot\cdot}, \widehat{\boldsymbol{\zeta}}^{(k)}]$

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

No passo M a maximização pode ser feita pelo algoritmo Newton-Raphson ou "Scoring"de Fisher. Com a suposição de que os itens são independentes, (3.80), a matriz de derivadas segundas torna-se bloco diagonal, possibilitando que os (parâmetros dos) itens sejam estimados individualmente, eliminando o problema de trabalhar com matrizes de ordem $3I \times 3I$ e passando a operar com matrizes 3×3.

Há três formas do algoritmo EM, distinguidas pela relação entre a função (densidade) de probabilidade e a forma da família exponencial. A primeira forma se aplica quando a função é um membro regular da família exponencial; a segunda, quando a função não é um membro regular da família exponencial, mas um membro da *família exponencial curvada* (formada por distribuições em que há restrições no espaço paramétrico) e a terceira, quando a função não tem nenhuma relação com a família exponencial.

Se a FRI é um membro regular da família exponencial, o procedimento torna-se relativamente simples. Embora o modelo logístico de 1 parâmetro (modelo de Rasch) seja membro da família exponencial, os modelos de 2 e 3 parâmetros não são. Portanto, a terceira forma do algoritmo EM deve ser aplicada nestes casos.

Para descrever brevemente o algoritmo EM aplicado à TRI, comecemos supondo que as habilidades estão restritas a um conjunto de q valores, $\overline{\theta}_k,\ k =$

$1, \cdots, q$, com probabilidades $\pi_k,\ k = 1, \cdots, q$. (Essa suposição pode ser feita porque as aproximações de integrais são feitas por métodos de quadratura, e os valores $\overline{\theta}_k$ corresponderão aos pontos de quadratura.) Seja f_{ki} o números de indivíduos com habilidade $\overline{\theta}_k$ respondendo ao item i, $\boldsymbol{f}_i = (f_{1i}, \cdots, f_{qi})'$, com $\sum_{k=1}^{q} f_{ki} = N$, $\boldsymbol{f} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_I)$. Similarmente, seja r_{ki} o números de indivíduos com habilidade $\overline{\theta}_k$ respondendo corretamente ao item i, $\boldsymbol{r}_i = (r_{1i}, \cdots, r_{qi})'$ e $\boldsymbol{r} = (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_I)$. Estas definições se assemelham bastante com as da Seção 3.2, quando tratamos da estimação dos parâmetros dos itens com as habilidades conhecidas e agrupadas em q categorias. Veremos que, de fato, os resultados são muito similares. Entretanto, na Seção 3.2 as freqüências f_{ki} e r_{ki} eram conhecidas, e no caso atual estas quantidades são desconhecidas. Essa é a grande vantagem do algoritmo EM, onde f_{ki} e r_{ki} podem ser tratadas como quantidades não observadas.

Se os n indivíduos que responderão ao item i são selecionados ao acaso da população, a probabilidade conjunta que os f_{ki} indivíduos tenham habilidades $\overline{\theta}_k$, $k = 1, \cdots, q$, é dada pela distribuição multinomial:

$$P(\boldsymbol{f}_i|\boldsymbol{\pi}) = \frac{n!}{\prod_{k=1}^{q} f_{ki}!} \prod_{k=1}^{q} \pi_j^{f_{ki}}, \quad i = 1, \cdots, I.$$

Dados f_{ki} e $\overline{\theta}_k$, a probabilidade de ocorrerem r_{ki} acertos ao item i dentre as f_{ki} tentativas por indivíduos com habilidade $\overline{\theta}_k$ é

$$P(r_{ki}|f_{ki}, \overline{\theta}_k) = \binom{f_{ki}}{r_{ki}} P_{ki}^{r_{ki}} Q_{ki}^{f_{ki}-r_{ki}},$$

onde P_{ki} é a FRI adotada com θ_j substituída por $\overline{\theta}_k$. A probabilidade conjunta de $\boldsymbol{f}$ e $\boldsymbol{r}$, dados $\overline{\boldsymbol{\theta}} = (\overline{\theta}_1, \cdots, \overline{\theta}_q)$ e $\boldsymbol{\pi}$, é

$$\begin{aligned} P(\boldsymbol{f}, \boldsymbol{r}|\overline{\boldsymbol{\theta}}, \boldsymbol{\pi}) &= P(\boldsymbol{f}|\overline{\boldsymbol{\theta}}, \boldsymbol{\pi}) P(\boldsymbol{r}|\boldsymbol{f}, \overline{\boldsymbol{\theta}}, \boldsymbol{\pi}) \\ &= P(\boldsymbol{f}|\boldsymbol{\pi}) P(\boldsymbol{r}|\boldsymbol{h}, \overline{\boldsymbol{\theta}}) \\ &= \left\{ \prod_{i=1}^{I} P(\boldsymbol{f}_i|\boldsymbol{\pi}) \right\} \left\{ \prod_{i=1}^{I} \prod_{k=1}^{q} P(r_{ki}|f_{ki}, \overline{\theta}_k) \right\} \end{aligned}$$

$$\begin{aligned}
\log L(\boldsymbol{\zeta}) &= \log P(\boldsymbol{f}|\boldsymbol{\pi}) + \sum_{i=1}^{I}\sum_{k=1}^{q} \log P(r_{ki}|f_{ki},\overline{\theta}_k) \\
&= \log P(\boldsymbol{f}|\boldsymbol{\pi}) + \sum_{i=1}^{I}\sum_{k=1}^{q} \left\{ \log \binom{f_{ki}}{r_{ki}} + r_{ki}\log P_{ki} + (f_{ki}-r_{ki})\log Q_{ki} \right\} \\
&= C + \sum_{k=1}^{q}\sum_{i=1}^{I} \left\{ r_{ki}\log P_{ki} + (f_{ki}-r_{ki})\log Q_{ki} \right\},
\end{aligned}$$

onde $C = \log P(\boldsymbol{f}|\boldsymbol{\pi}) + \sum_{i=1}^{I}\sum_{k=1}^{q} \log \binom{f_{ki}}{r_{ki}}$ é constante com relação a $\boldsymbol{\zeta}$. Temos que $(\boldsymbol{f},\boldsymbol{r})$ são não-observáveis, mas tomando a esperança da log-verossimilhança, condicional em $\boldsymbol{u}_{..}$ e $\boldsymbol{\zeta}$, e usando a notação

$$\overline{r}_{ki} = E[r_{ki}|\boldsymbol{u}_{..},\boldsymbol{\zeta}], \quad \overline{f}_{ki} = E[f_{ki}|\boldsymbol{u}_{..},\boldsymbol{\zeta}] \quad \text{e} \quad \overline{C} = E[C|\boldsymbol{u}_{..},\boldsymbol{\zeta}]$$

obtemos,

$$E[\log L(\boldsymbol{\zeta})] = \overline{C} + \sum_{i=1}^{I}\sum_{k=1}^{q} \left\{ \overline{r}_{ki}\log P_{ki} + (\overline{f}_{ki}-\overline{r}_{ki})\log Q_{ki} \right\}. \quad (3.88)$$

Podemos notar que esta expressão equivale a (3.17) da Seção 3.2. As primeiras parcelas nessas duas expressões são constantes com relação a $\boldsymbol{\zeta}$. Os termos restantes são, praticamente, os mesmos, com f_{ki} e r_{ki} substituídos por $\overline{f}_{ki}$ e $\overline{r}_{ki}$, respectivamente. Portanto, maximizar a equação (3.88) com relação a $\boldsymbol{\zeta}_i$ é equivalente a maximizar (3.17) e representa o Passo E do algoritmo EM. Mais especificamente, os passos E e M são

Passo E Usar os pontos de quadratura $\overline{\theta}_k$, os pesos A_k, $k = 1,\cdots,q$ e estimativas iniciais dos parâmetros dos itens, $\widehat{\boldsymbol{\zeta}}_i$, $i = 1,\cdots,I$, para gerar $g_j^*(\overline{\theta}_k)$ e, posteriormente, $\overline{r}_{ki}$ e $\overline{f}_{ki}$, $i = 1,\cdots,I$ e $k = 1,\cdots,q$.

Passo M Com $\boldsymbol{r}$ e $\boldsymbol{f}$ obtidos no Passo E, resolver as equações de estimação para $\boldsymbol{\zeta}_i$, $i = 1,\cdots,I$, usando o algoritmo Newton-Raphson ou "Scoring"de Fisher através das expressões da Seção 3.2.

Estes passos compõem cada iteração do algoritmo EM, as quais serão repetidas até que algum critério de parada seja alcançado. Após a finalização do processo, os erros-padrão são obtidos com o uso de (3.29).

3.6 Estimação bayesiana

A estimação por máxima verossimilhança apresenta problemas na estimação de itens que são respondidos corretamente, ou incorretamente, por todos os indivíduos, e também das habilidades de indivíduos que responderam corretamente, ou incorretamente, a todos os itens. Além disso, há a possibilidade de que as estimativas dos parâmetros dos itens caiam fora do intervalo esperado, tal como valores de a_i negativos, ou valores de c_i fora do intervalo $[0, 1]$. A metodologia bayesiana apresenta uma solução em que estes problemas são contornados.

Há várias propostas para a estimação bayesiana dos parâmetros de interesse na TRI. A mais utilizada é a *Estimação Bayesiana Marginal* proposta por Mislevy (1986a), que é uma generalização da proposta de Bock & Aitkin (1981). Basicamente, a estimação bayesiana consiste em estabelecer distribuições *a priori* para os parâmetros de interesse, construir uma nova função denominada distribuição *a posteriori* e estimar os parâmetros de interesse com base em alguma característica dessa distribuição. Consideremos que as componentes de $\boldsymbol{\zeta}$ são variáveis aleatórias independentes e contínuas, com distribuições especificadas. Por estarmos tratando de uma extensão da proposta de Bock & Aitkin, a estimação é feita por máxima verossimilhança marginal, em duas etapas.

Para tornar o tratamento mais geral, vamos considerar que a distribuição da habilidade é função de um vetor de parâmetros $\boldsymbol{\eta}$, com densidade $g(\theta|\boldsymbol{\eta})$, e que a distribuição de $\boldsymbol{\zeta}_i$, $i = 1, \cdots, I$, é função de um vetor de parâmetros $\boldsymbol{\tau}$, com densidade $g(\boldsymbol{\zeta}|\boldsymbol{\tau})$. Podemos, ainda, estabelecer distribuições *a priori* para os parâmetros $\boldsymbol{\tau}$ e $\boldsymbol{\eta}$, digamos $f(\boldsymbol{\tau})$ e $g(\boldsymbol{\eta})$. Para definir a métrica, digamos (0,1), em que os parâmetros dos itens (e, posteriormente, as habilidades) serão estimados, podemos adotar uma distribuição degenerada para $\boldsymbol{\eta}$ em (0,1) ou uma distribuição que tenha vetor de médias (0,1) e variâncias muito pequenas. A primeira opção equivale a eliminar a função $g(\boldsymbol{\eta})$, mas para tornar o trata-

mento mais geral vamos mantê-la no desenvolvimento da teoria. Com isso, a densidade conjunta desses parâmetros é

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\tau}) &= f(\boldsymbol{\zeta}|\boldsymbol{\tau}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) f(\boldsymbol{\tau}) g(\boldsymbol{\eta}) \\ &= \left\{ \prod_{i=1}^{I} f(\boldsymbol{\zeta}_i|\boldsymbol{\tau}) \right\} \left\{ \prod_{j=1}^{n} g(\theta_j|\boldsymbol{\eta}) \right\} f(\boldsymbol{\tau}) g(\boldsymbol{\eta}). \end{aligned}$$

Se quisermos fazer inferências sobre todos esses parâmetros, devemos nos basear na distribuição a posteriori:

$$f(\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\tau}|\boldsymbol{u}_{..}) \propto L(\boldsymbol{u}_{..}; \boldsymbol{\theta}, \boldsymbol{\zeta}) f(\boldsymbol{\zeta}|\boldsymbol{\tau}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) f(\boldsymbol{\tau}) g(\boldsymbol{\eta}). \tag{3.89}$$

Entretanto, geralmente estamos interessados em um número reduzido de parâmetros. Nesse caso, devemos trabalhar com uma posteriori que seja função apenas dos parâmetros de interesse.

3.6.1 Estimação dos parâmetros dos itens

Para fazer inferências com relação aos parâmetros dos itens, é conveniente "marginalizar"a posteriori integrando com relação a θ e $\boldsymbol{\tau}$, obtendo a distribuição *a posteriori* de $\boldsymbol{\zeta}$ e $\boldsymbol{\eta}$:

$$\begin{aligned} f^*(\boldsymbol{\zeta}, \boldsymbol{\eta}|\boldsymbol{u}_{..}) &= C \int \int P(\boldsymbol{u}_{..}; \boldsymbol{\theta}, \boldsymbol{\zeta}) f(\boldsymbol{\zeta}|\boldsymbol{\tau}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) f(\boldsymbol{\tau}) g(\boldsymbol{\eta}) d\theta d\boldsymbol{\tau} \\ &= C g(\boldsymbol{\eta}) \left\{ \int f(\boldsymbol{\zeta}|\boldsymbol{\tau}) f(\boldsymbol{\tau}) d\boldsymbol{\tau} \right\} \left\{ \int P(\boldsymbol{u}_{..}; \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\theta \right\} \\ &\propto L(\boldsymbol{\zeta}, \boldsymbol{\eta}) f(\boldsymbol{\zeta}) g(\boldsymbol{\eta}), \end{aligned} \tag{3.90}$$

onde C representa uma constante, $L(\boldsymbol{\zeta}, \boldsymbol{\eta}) \equiv P(\boldsymbol{u}_{..}; \boldsymbol{\zeta}, \boldsymbol{\eta})$ e

$$f(\boldsymbol{\zeta}) = \int f(\boldsymbol{\zeta}|\boldsymbol{\tau}) f(\boldsymbol{\tau}) d\boldsymbol{\tau}.$$

Como estimador de $\boldsymbol{\zeta}$ podemos escolher alguma característica de $f^*(\boldsymbol{\zeta}, \boldsymbol{\eta}|\boldsymbol{u}_{..})$,

sendo que as mais adotadas são a média e a moda. No que segue vamos considerar a *moda da posteriori* como o estimador de $\boldsymbol{\zeta}$, ou seja, o valor de $\boldsymbol{\zeta}$ que maximiza a posteriori marginal. Temos que

$$\log f^*(\boldsymbol{\zeta}, \boldsymbol{\eta}|\boldsymbol{u}_{..}) = Const + \log L(\boldsymbol{\zeta}, \boldsymbol{\eta}) + \log f(\boldsymbol{\zeta}) + \log g(\boldsymbol{\eta}),$$

onde o primeiro termo representa uma constante. Pela suposição de independência entre os itens, a estimação será feita um item por vez. Notando que a última parcela não é função de $\boldsymbol{\zeta}_i$, temos que as equações de estimação para os parâmetros dos itens $\boldsymbol{\zeta}_i,\ i = 1, \cdots, I$, são dadas por

$$\frac{\partial f^*(\boldsymbol{\zeta}, \boldsymbol{\eta}|\boldsymbol{u}_{..})}{\partial \boldsymbol{\zeta}_i} = \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} + \frac{\partial \log f(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} = \mathbf{0}. \tag{3.91}$$

A primeira parcela de (3.91) é exatamente a mesma obtida em (3.64). A abordagem bayesiana adiciona uma nova parcela a (3.64) relativa à distribuição *a priori* associada aos parâmetros dos itens. A primeira parcela de (3.91) relativa às componentes de $\boldsymbol{\zeta}_i$ é dada por (3.68) a (3.70). A segunda parcela de (3.91) depende da priori adotada para cada parâmetro. Como o parâmetro a_i deve ser positivo, b_i pode assumir qualquer valor real e c_i deve estar no intervalo $[0, 1]$, deveremos assumir distribuições que levam em conta essas limitações e isso exige um tratamento diferenciado para cada um destes parâmetros. Em seguida trataremos destes casos, considerando as suposições mais freqüentes na prática.

Distribuição a priori para a_i

Geralmente, adota-se as distribuições *Log-normal* ou *Chi-Quadrado* para a_i. Neste texto, vamos supor que cada parâmetro a_i tem distribuição Log-normal com parâmetro $\boldsymbol{\tau} = (\mu_a, \sigma_a^2)$. Uma justificativa teórica para a adoção desta distribuição é que na prática os a_i são, em geral, positivos, sugerindo que a distribuição de a_i pode ser modelada por uma distribuição unimodal e com assimetria positiva (ver Mislevy (1986a)), tal como a log-normal. A transformação $\alpha_i = \log a_i$ resulta em cada α_i tendo uma distribição Normal $(\mu_\alpha, \sigma_\alpha^2)$, onde $\mu_a = \exp[\mu_\alpha + \sigma_\alpha^2/2]$ e $\sigma_a^2 = (\exp(\sigma_\alpha^2) - 1)\exp[2\mu_\alpha + \sigma_\alpha^2]$. Alguns autores (ver Baker (1992), por exemplo) preferem desenvolver expressões para

a estimação de α_i ao invés de a_i e sugerem a utilização da propriedade de invariância do estimador de máxima verossimilhança para a obtenção de $\widehat{a}_i$ pela trasformação $\widehat{a}_i = \exp(\widehat{\alpha}_i)$. Entretanto, para uniformidade desse texto, vamos continuar apresentando a equação para o parâmetro a_i.

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

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

Segue que a segunda parcela de (3.91) pode ser escrita como

$$\frac{\partial \log f(a_i|\mu_a, \sigma_a^2)}{\partial a_i} = -\frac{1}{a_i}\left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2}\right]. \tag{3.92}$$

Distribuição a priori para b_i

Como os parâmetros de dificuldade estão na mesma escala da habilidade, em geral, supõem-se que cada b_i´s tem distribuição *Normal* com vetor de parâmetros $\boldsymbol{\tau} = (\mu_b, \sigma_b^2)$. Desta forma, a segunda parcela de (3.91) pode ser escrita como

$$\frac{\partial \log f(b_i|\mu_b, \sigma_b^2)}{\partial b_i} = -\frac{(b_i - \mu_b)}{\sigma_b^2}. \tag{3.93}$$

Distribuição a priori para c_i

Como c_i só pode pertencer ao intervalo $[0; 1]$, uma priori *Beta* foi proposta por Swaminathan & Gifford (1986). A função densidade da distribuição Beta com parâmetros $s + 1$ e $t + 1$ é dada por

$$f(c_i|s, t) = \frac{\Gamma(s + t + 2)}{\Gamma(s + 1)\Gamma(t + 1)} c_i^s (1 - c_i)^t, \tag{3.94}$$

onde $\Gamma(d)$ é a função Gama, definida por

$$\Gamma(d) = \int_0^\infty x^{d-1} e^{-x} dx.$$

A média desta distribuição é dada por

$$p = \frac{s+1}{s+t+2}.$$

Swaminathan & Gifford propõem, ainda, a seguinte reparametrização:

$$\alpha = mp + 1 \quad \text{e} \quad \beta = m(1-p) + 1,$$

onde $m = s+t+2$. Desta forma, $p = (s+1)/m$ e, consequentemente, $s = mp-1$ e $t = m - s - 2 = m(1-p) - 1$. Segue disso que

$$s = \alpha - 2 \quad \text{e} \quad t = \beta - 2.$$

Retornando a (3.94), obtemos

$$f(c_i|\alpha, \beta) = \frac{\Gamma(\alpha+\beta-2)}{\Gamma(\alpha-1)\Gamma(\beta-1)} c_i^{\alpha-2}(1-c_i)^{\beta-2}. \tag{3.95}$$

Neste caso, a média p passa a ser interpretada como a probabilidade de acerto por indivíduos com baixa habilidade. Desta forma, os parâmetros α e β são definidos para que p tenha o valor desejado. Entretanto, Swaminathan & Gifford sugerem que a escolha de m deva se situar no intervalo $\{15, \cdots, 20\}$, o que leva a uma certa restrição na escolha de α e β.

Para chegarmos a expressão para a segunda parcela de (3.91), notemos que

$$\log f(c_i|\alpha, \beta) = Const + (\alpha - 2)\log c_i + (\beta - 2)\log(1 - c_i). \tag{3.96}$$

Consequentemente,

$$\frac{\partial \log f(c_i|\alpha, \beta)}{\partial c_i} = \frac{\alpha - 2}{c_i} - \frac{\beta - 2}{1 - c_i}. \tag{3.97}$$

Com as componentes (3.92), (3.93) e (3.97), temos que as equações de estimação para as componentes de $\boldsymbol{\zeta}_i$ são

$$a_i \;:\; D(1-c_i)\int_{I\!R}(\theta-b_i)\left[r_i(\theta)-P_i f_i(\theta)\right]W_i d\theta - \frac{1}{a_i}\left[1+\frac{\log a_i-\mu_a}{\sigma_a^2}\right]=0, \tag{3.98}$$

$$b_i \;:\; -Da_i(1-c_i)\int_{I\!R}\left[r_i(\theta)-P_i f_i(\theta)\right]W_i d\theta - \frac{(b_i-\mu_b)}{\sigma_b^2}=0, \tag{3.99}$$

$$c_i \;:\; \int_{I\!R}\left[r_i(\theta)-P_i f_i(\theta)\right]\frac{W_i}{P_i^*}d\theta + \frac{\alpha-2}{c_i} - \frac{\beta-2}{1-c_i}=0. \tag{3.100}$$

Para efeito de aplicação dos procedimentos iterativos Newton-Raphson ou "Scoring"de Fisher, precisaremos das derivadas segundas das expressões (3.98) a (3.100). Como as derivadas segundas das primeiras parcelas dessas expressões já foram obtidas na Seção 3.2, resta apenas a obtenção das segundas parcelas, que são as seguintes:

$$\begin{aligned}
\frac{\partial^2 \log f(a_i|\mu_a,\sigma_a^2)}{\partial a_i^2} &= \frac{1}{a_i\sigma_a^2}\left[\sigma_a^2+\log a_i-\mu_a-1\right], \\
\frac{\partial^2 \log f(b_i|\mu_b,\sigma_b^2)}{\partial b_i^2} &= -\frac{1}{\sigma_b^2}, \\
\frac{\partial^2 \log f(c_i|\alpha,\beta)}{\partial c_i^2} &= -\frac{\alpha-2}{c_i^2}-\frac{\beta-2}{(1-c_i)^2}.
\end{aligned} \tag{3.101}$$

Equações de estimação em forma de quadratura

Considerando conhecidos os nós $\overline{\theta}_k$ e os pesos A_k, $k=1,\cdots,q$, temos que as equações de estimação em forma de quadratura para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i \;:\; D(1-c_i)\sum_{k=1}^{q}(\overline{\theta}_k - b_i)\left[r_{ki} - P_{ki}f_{ki}\right]W_{ki} - \frac{1}{a_i}\left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2}\right] = 0, \tag{3.102}$$

$$b_i \;:\; -Da_i(1-c_i)\sum_{k=1}^{q}\left[r_{ki} - P_{ki}f_{ki}\right]W_{ki} - \frac{(b_i - \mu_b)}{\sigma_b^2} = 0, \tag{3.103}$$

$$c_i \;:\; \sum_{k=1}^{q}\left[r_{ki} - P_{ki}f_{ki}\right]\frac{W_{ki}}{P^*_{ki}} + \frac{\alpha - 2}{c_i} - \frac{\beta - 2}{1 - c_i} = 0. \tag{3.104}$$

3.6.2 Estimação das habilidades

Tal como na estimação por máxima verossimilhança marginal, a estimação bayesiana das habilidades é feita em uma segunda etapa, considerando os parâmetros dos itens fixos. Através da suposição de independência entre as habilidades de diferentes indivíduos, podemos fazer as estimações em separado para cada indivíduo.

Vamos assumir que a distribuição a priori para $\theta_j,\ j = 1, \cdots, n$, é Normal com vetor de parâmetros $\boldsymbol{\eta} = (\mu, \sigma^2)$ conhecidos. A posteriori para a habilidade do indivíduo j pode ser escrita como

$$g_j^*(\theta_j) \equiv g(\theta_j|\boldsymbol{u}_{j.}, \boldsymbol{\zeta}, \boldsymbol{\eta}) \propto P(\boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta})g(\theta_j|\boldsymbol{\eta}). \tag{3.105}$$

Novamente, podemos adotar alguma característica de $g_j^*(\theta_j)$ como estimador de θ_j, sendo que as mais adotadas são a média e a moda. A seguir, trataremos da obtenção de cada uma destas características.

Estimação pela moda da posteriori - MAP

A estimação pela moda da posteriori (ou MAP: maximum a posteriori) consiste em obter o máximo de (3.105). Por facilidade, vamos trabalhar com o logaritimo da posteriori

$$\log g_j^*(\theta_j) = Const + \log P(\boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta}) + \log g(\theta_j|\boldsymbol{\eta}).$$

Segue que a equação de estimação para θ_j é

$$\frac{\partial \log g_j^*(\theta_j)}{\partial \theta_j} = \frac{\partial \log P(\boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta})}{\partial \theta_j} + \frac{\partial \log g(\theta_j|\boldsymbol{\eta})}{\partial \theta_j} = 0. \tag{3.106}$$

Pela independência local, temos que

$$\log P(\boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta}) = \log \left[\prod_{i=1}^{I} P(u_{ji}|\boldsymbol{\zeta}_i, \theta_j) \right] = \sum_{i=1}^{I} \log P(u_{ji}|\boldsymbol{\zeta}_i, \theta_j).$$

Portanto,

$$\begin{aligned} \frac{\partial \log P(\boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta})}{\partial \theta_j} &= \sum_{i=1}^{I} \frac{\partial \log P(u_{ji}|\boldsymbol{\zeta}_i, \theta_j)}{\partial \theta_j} \\ &= \sum_{i=1}^{I} \frac{\partial P(u_{ji}|\boldsymbol{\zeta}_i, \theta_j)/\partial \theta_j}{P(u_{ji}|\boldsymbol{\zeta}_i, \theta_j)}. \end{aligned} \tag{3.107}$$

Lembramdo que $P(u_{ji}|\boldsymbol{\zeta}_i, \theta_j) = P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}}$ e usando o desenvolvimento de (3.34) a (3.38), teremos que

$$\frac{\partial \log P(\boldsymbol{u}_{j.}|\theta_j, \boldsymbol{\zeta})}{\partial \theta_j} = D \sum_{i=1}^{I} a_i(1 - c_i)(u_{ji} - P_{ji})W_{ji}. \tag{3.108}$$

Como estamos adotando a priori Normal (μ, σ^2) para θ_j, a segunda parcela de (3.106) é

$$\frac{\partial \log g(\theta_j|\boldsymbol{\eta})}{\partial \theta_j} = -\frac{(\theta_j - \mu)}{\sigma^2}. \tag{3.109}$$

Por (3.108) e (3.109), temos que a equação de estimação para θ_j é

$$\theta_j \quad : \quad D \sum_{i=1}^{I} a_i(1 - c_i)(u_{ji} - P_{ji})W_{ji} - \frac{(\theta_j - \mu)}{\sigma^2} = 0. \tag{3.110}$$

Como esta equação não tem solução explícita, podemos aplicar algum método iterativo para resolvê-la. Para isso será necessária a derivada segunda de $\log g(\theta_j|\boldsymbol{u}_{j.}, \boldsymbol{\zeta}, \boldsymbol{\eta})$ com relação a θ_j, cuja expressão é

$$H(\theta_j) = \sum_{i=1}^{I}(u_{ji} - P_{ji})W_{ji}\left\{H_{ji} - (u_{ji} - P_{ji})W_{ji}h_{ji}^2\right\} - \frac{1}{\sigma^2}, \qquad (3.111)$$

onde h_{ji} e H_{ji} são dados por (3.42) e (3.43), respectivamente. Para aplicarmos o método "Scoring"de Fisher, devemos tomar a esperança da expressão acima, resultando em

$$\Delta(\theta_j) = -\sum_{i=1}^{I} P_{ji}^* Q_{ji}^* W_{ji} h_{ji}^2 - \frac{1}{\sigma^2}. \qquad (3.112)$$

Estimação pela média da posteriori - EAP

A estimação de θ_j pela média da posteriori (ou EAP: expected a posteriori) consiste em obter a esperança da posteriori, que pode ser escrita como

$$g(\theta|\boldsymbol{u}_{j.}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \frac{P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta})}. \qquad (3.113)$$

Segue que a esperança da posteriori é

$$\widehat{\theta}_j \equiv E[\theta|\boldsymbol{u}_{j.}, \boldsymbol{\zeta}, \boldsymbol{\eta}] = \frac{\int_{I\!R} \theta g(\theta|\boldsymbol{\eta})P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta})d\theta}{\int_{I\!R} g(\theta|\boldsymbol{\eta})P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta})d\theta}. \qquad (3.114)$$

Esta forma de estimação tem a vantagem de ser calculada diretamente, não necessitando da aplicação de métodos iterativos. Além disso, as quantidades necessárias para o seu cálculo são um produto final da etapa de estimação. Por conta disso alguns autores (por exemplo, Mislevy & Stocking (1989)) recomendam esta escolha para a estimação das habilidades.

3.7 Resumo

A seguir, faremos um síntese das vantagens e desvantagens dos métodos citados neste livro. Vale ressaltar que existem ainda outros métodos de estimação propostos na literatura.

Na síntese abaixo, o símbolo $\oplus$ representará uma característica positiva, enquanto $\ominus$ representará uma característica negativa.

Estimação dos Parâmetros dos Itens

- **Máxima Verossimilhança Marginal - MVM** :

  $\oplus$ Possui propriedades assintóticas: as estimativas dos parâmetros a_i, b_i e c_i são consistentes;

  $\oplus$ Uma vez estimados os parâmetros dos itens, pode-se estimar as habilidades através do método da Seção 3.3 ou 3.6.2;

  $\ominus$ Não está definido para itens com acerto total ou erro total;

  $\ominus$ É bastante trabalhoso computacionalmente;

  $\ominus$ Necessidade do estabelecimento de uma distribuição para θ;

  $\ominus$ Apresenta problemas na estimação do parâmetro c_i em alguns casos; deve ser usado somente com um número suficientemente grande de respondentes.

- **Bayesiano** :

  $\oplus$ Definido para qualquer padrão de resposta;

  $\oplus$ Uma vez estimados os parâmetros dos itens, pode-se estimar as habilidades através do método da Seção 3.3 ou 3.6.2;

  $\ominus$ É mais trabalhoso computacionalmente do que o MVM;

  $\ominus$ Necessidade de distribuições a priori para os parâmetros dos itens.

Estimação das habilidades

- **Máxima Verossimilhança - MV** :
  - ⊕ Para testes "longos"produz estimadores não viciados;
  - ⊖ Não está definido para alguns padrões de resposta.
- **Bayesiano - EAP** :
  - ⊕ Definido para qualquer padrão de resposta;
  - ⊕ Possui o menor erro médio;
  - ⊖ Viciado;
  - ⊖ Exige cálculos mais complexos do que o método de MV;
  - ⊖ Necessidade de uma distribuição a priori para θ.
- **Bayesiano - MAP** :
  - ⊕ Definido para qualquer padrão de resposta;
  - ⊖ Viciado.
  - ⊖ Exige cálculos mais complexos do que o método de MV;
  - ⊖ Necessidade de uma distribuição a priori para θ.

Estimação dos parâmetros dos itens e das habilidades

- **Máxima Verossimilhança Conjunta - MVC** :
  - ⊕ Serve como base para outros procedimentos;
  - ⊖ Apresenta problemas de indeterminação;
  - ⊖ Não possui propriedades assintóticas, pois o aumento do número de respondentes aumenta o número de parâmetros a serem estimados;
  - ⊖ É bastante trabalhoso computacionalmente;

- ⊖ Apresenta problemas na estimação do parâmetro c_i em alguns casos; deve ser usado somente com um número suficientemente grande de respondentes;
- ⊖ Não está definido para alguns padrões de resposta.

No próximo capítulo introduziremos e discutiremos o conceito de equalização.

Capítulo **4**

Equalização

4.1 Introdução

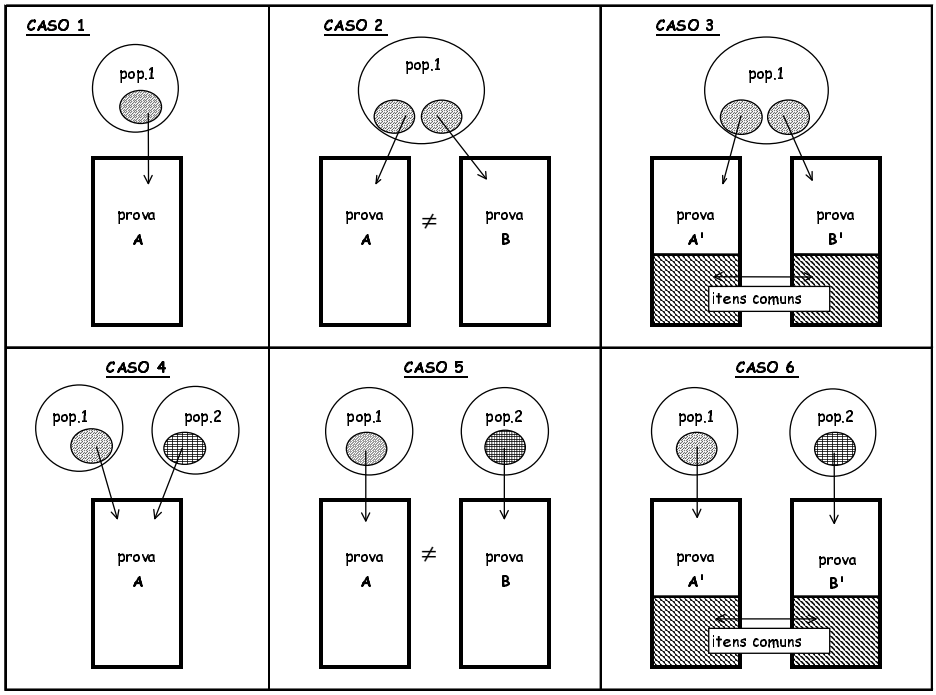
No capítulo anterior, apresentamos os métodos de estimação mais utilizados quando *todos* os parâmetros dos itens de uma *única* prova devem ser estimados. No entanto, esta é apenas uma das possíveis situações que na prática iremos encontrar. A seguir, listaremos os 6 casos possíveis, quanto ao número de grupos e de tipos de prova envolvidos. Esses casos estão esquematizados na Figura 4.1.

1. Um único grupo fazendo uma única prova.
2. Um único grupo, dividido em dois subgrupos, fazendo duas provas, totalmente distintas (nenhum item comum).
3. Um único grupo, dividido em dois subgrupos, fazendo duas provas, apenas parcialmente distintas, ou seja, com alguns itens comuns.
4. Dois grupos fazendo uma única prova.
5. Dois grupos fazendo duas provas, totalmente distintas (nenhum item comum).
6. Dois grupos fazendo duas provas, apenas parcialmente distintas, ou seja, com alguns itens comuns.

Note que para simplificar, listamos os casos acima utilizando apenas duas provas e duas populações, mas as situações envolvendo um número maior de provas e/ou de populações são análogas.

Além disso, os problemas de estimação também podem diferir dependendo do conjunto de itens que necessita ser estimado, ou seja, se nosso conjunto de itens é composto de:

Figura 4.1 *Representação gráfica de 6 situações quanto ao número de grupos e de tipos de provas*



(a) apenas itens novos (ou seja, itens que ainda não foram calibrados);

(b) apenas itens já calibrados;

(c) itens novos e itens calibrados.

Em primeiro lugar, é importante definir o conceito de *Equalização* (ver Kolen & Brennan (1995), por exemplo), que é um dos mais importantes da TRI e um dos grandes objetivos das Avaliações Educacionais. Equalizar significa equiparar, tornar comparável, o que no caso da TRI significa colocar parâmetros de itens vindos de provas distintas ou habilidades de respondentes de diferentes

grupos, na mesma métrica, isto é, numa escala comum, tornando os itens e/ou as habilidades comparáveis.

Existem dois tipos de equalização: a equalização *via população* e a equalização *via itens comuns*. Isto significa que há duas maneiras de colocar parâmetros, tanto de itens quanto de habilidades, numa mesma métrica: na primeira usamos o fato de que se um único grupo de respondentes é submetido a provas distintas, basta que todos os itens sejam calibrados conjuntamente para termos a garantia de que todos estarão na mesma métrica. Já na equalização via itens comuns, a garantia de que as populações envolvidas terão seus parâmetros em uma única escala será dada pelos itens comuns entre as populações, que servirão de ligação entre elas.

4.2 Diferentes tipos de equalização

Uma vez listadas as diversas situações e casos que podemos ter, vamos agora discutir cada uma delas. Obviamente, podemos ter as situações 1 a 6 combinadas com os casos (a) a (c). Mas, mais uma vez para facilitar a explicação, trataremos inicialmente das situações 1 a 6 considerando sempre o caso mais simples, ou seja, o caso (a).

Cabe ainda ressaltar que todas as análises e comentários deste capítulo serão feitos considerando-se o modelo logístico unidimensional de 3 parâmetros.

4.2.1 Um único grupo fazendo uma única prova

Este é o caso trivial, em que se aplicam diretamente os modelos matemáticos e os métodos de estimação descritos nos capítulos anteriores. Foi o caso considerado até agora, nos Capítulos 2 e 3, e pela própria natureza do problema, não é necessário nenhum tipo de equalização.

Um exemplo para ilustrar este caso seria uma prova de 30 itens aplicada à 4.^{a} série diurna do Ensino Fundamental da rede pública estadual de São Paulo.

4.2.2 Um único grupo fazendo duas provas totalmente distintas

Este é um caso clássico do que chamamos de equalização via população. Para resolvê-lo, basta que todos os itens de ambas as provas sejam calibrados

simultaneamente. O fato de todos os indivíduos representarem uma amostra aleatória de uma mesma população é que garante que todos os parâmetros envolvidos estarão na mesma escala.

Um exemplo para este caso seria quando duas provas distintas (tipo A e tipo B), com 30 itens cada, são aplicadas, de maneira aleatória, aos alunos da 4.ª série diurna do Ensino Fundamental da rede pública estadual de São Paulo. Ao final dos processos de estimação, todos os resultados obtidos serão comparáveis, não importando a que tipo de prova cada aluno tenha sido submetido.

4.2.3 Um único grupo fazendo duas provas parcialmente distintas

Este caso é bastante semelhante ao caso anterior, e aqui também podemos fazer a equalização via população. Assim, valem os mesmos comentários da Seção 4.2.2.

Um exemplo dessa situação seria a aplicação de duas provas (tipo A e tipo B), com 30 itens cada e com 10 itens comuns entre elas, aos alunos da 4.ª série diurna do Ensino Fundamental da rede pública estadual de São Paulo. Aqui, o número total de itens a serem calibrados seria 50 (= 30 + 30 - 10). Analogamente ao exemplo anterior, ao final dos processos de estimação todos os resultados obtidos serão comparáveis, não importando a qual prova cada aluno tenha sido submetido.

Outro exemplo bastante interessante para este caso, seria a aplicação do SAEB — Sistema Nacional de Avaliação da Educação Básica. Nesse estudo, uma das populações alvo é a 3.ª série do Ensino Médio. Como a aplicação é de caráter nacional, alunos de vários estados do país são avaliados, mas todos são considerados como respondentes vindos da mesma população, ou seja, como um único grupo. Além disso, o SAEB procura cobrir a grade curricular de forma completa, e para tanto, é considerado um grande número de itens distintos em cada disciplina. Como seria inviável a aplicação de todos os itens a um único aluno, as provas são montadas segundo um esquema BIB — Blocos Incompletos Balanceados — no qual os itens são divididos em blocos, que por sua vez são reunidos em cadernos, e estes cadernos — que nada mais são do que diferentes provas — , é que são aplicados aos alunos. No caso da 3.ª série do Ensino Médio, os itens foram divididos em 13 blocos com 13 itens distintos

cada um. Foram então montados 26 cadernos, cada um composto por 3 blocos distintos. Assim, cada aluno responde a 39 itens. É importante notar que diferentes blocos não têm itens comuns entre si, mas que diferentes cadernos podem — ou não — ter itens comuns: basta que tenham algum bloco em comum. Concluindo, desta maneira foram aplicados diferentes tipos de provas — representados pelos 26 cadernos — com itens comuns a um único grupo de respondentes — alunos da 3.ª série do Ensino Médio brasileiro.

O SAEB também é um bom exemplo prático do que chamamos de provas com *itens não apresentados*. Podemos considerar que a prova é composta dos 169 itens, mas que apenas 39 são submetidos a cada aluno. Consequentemente, temos 130 itens que não foram apresentados para cada aluno. Quando temos provas com um número originalmente grande de itens, podemos resolver o problema utilizando esquemas semelhantes ao usado no SAEB. Assim, o que inicialmente poderia ser considerado como uma única prova, pode vir a ser considerado como várias provas, se não submetermos todos os itens a todos os alunos.

4.2.4 Dois grupos fazendo uma única prova

Este é um exemplo de equalização via itens comuns (só que no caso, todos). Como as duas populações fazem exatamente a mesma prova, basta que os itens sejam calibrados utilizando-se as respostas dos respondentes de ambos os grupos *simultaneamente*. Para tanto, devemos apenas utilizar um modelo para duas populações, como apresentado no Capítulo 2. Detalhes sobre o processo de estimação serão fornecidos no Capítulo 5.

Um exemplo para este caso seria a aplicação de uma única prova, composta de 40 itens, nos períodos diurno (população 1) e noturno (população 2) da 8.ª série do Ensino Fundamental da rede pública estadual de São Paulo. Ao final dos processos de estimação, todos os resultados obtidos serão comparáveis, não importando a que população o aluno pertence.

4.2.5 Dois grupos fazendo duas provas totalmente distintas

Este é o único dos 6 casos que não pode ser resolvido pela TRI. Obviamente é possível calibrar separadamente os itens das duas provas, mas o problema é que *não* podemos fazer nenhum tipo de comparação entre os resultados ob-

tidos, uma vez que eles estarão em métricas *diferentes*. Neste caso, não faz sentido comparar os resultados destes dois grupos, assim como não faz sentido comparar diretamente 40^oC com 40^oF . Assim como essas duas temperaturas estão em escalas de medida diferentes, os parâmetros obtidos nestas duas provas também estarão. A diferença é que, no caso das temperaturas, há uma relação conhecida entre as duas escalas, e assim, é possível colocarmos uma das temperaturas na mesma escala que a outra, possibilitando então, a comparação. Já no caso das provas, não existe nenhuma relação entre elas e nem entre os dois grupos, que torne possível a comparação.

Um exemplo que ilustra esta situação seria a elaboração de duas provas distintas: uma, composta de 30 itens, seria aplicada à $4.^a$ série diurna (população 1) e a outra prova, composta de 40 itens, seria aplicada à $5.^a$ série diurna (população 2) do Ensino Fundamental da rede pública estadual de São Paulo. Estas duas provas poderiam ser calibradas separadamente e seus resultados poderiam ser interpretados isoladamente dentro de cada série, mas não poderíamos comparar os resultados dos itens e nem das habilidades estimadas para os indivíduos das duas séries.

4.2.6 Dois grupos fazendo duas provas parcialmente distintas

Finalmente, vamos comentar o caso em que dois grupos são submetidos a duas provas diferentes, mas que têm alguns itens comuns. Assim como na Seção 4.2.4, este também é um exemplo de equalização via itens comuns. Este caso representa o melhor exemplo do uso e da importância da equalização e sem dúvida, ilustra o maior avanço da TRI sobre a Teoria Clássica. O uso de itens comuns entre provas distintas aplicadas a populações distintas permite que todos os parâmetros estejam na mesma escala ao final dos processos de estimação, possibilitando comparações e a construção de "escalas do conhecimento"interpretáveis, que são de grande importância na área educacional. A resolução deste caso é bastante semelhante ao que foi descrito na Seção 4.2.4, com a diferença que aqui apenas alguns dos itens (e não a prova toda) fazem a ligação entre as duas populações envolvidas. Este caso será abordado mais detalhadamente através de um exemplo prático, apresentado no Capítulo 6.

Um exemplo que ilustra bem esta situação seria a aplicação de uma prova com 30 itens à $3.^a$ série diurna (população 1) e de outra prova, também com 30 itens, à $4.^a$ série diurna da rede pública estadual de São Paulo (população

2). Entre elas poderiam haver 10 itens comuns (por exemplo, 10 itens da matriz curricular da 3.ª série). Desta maneira, no final do processo de estimação teríamos todos os 50 itens numa mesma métrica, possibilitando comparações entre alunos de 3.ª e 4.ª séries, e também possibilitando a criação de uma "escala de conhecimento"da 3.ª e da 4.ª série nesta dada disciplina. Como veremos no Capítulo 6, esta escala possibilitaria a verificação dos conteúdos que os alunos destas duas séries dominam, dos conteúdos onde há falhas, acompanhar a "evolução do conhecimento"de uma série para outra, etc.

4.3 Diferentes problemas de estimação

Vamos agora considerar outro ponto bastante importante na TRI: o conjunto de itens a ser calibrado. Vamos comentar inicialmente os casos (a) a (c), considerando-se o caso 1, ou seja, o caso em que uma única prova foi aplicada a um único grupo de respondentes.

4.3.1 Quando todos os itens são novos

Neste caso, todos os itens são considerados "novos", ou seja, deseja-se calibrar o conjunto completo de itens. Este é o caso trivial, que foi considerado até agora. Para resolver este problema basta utilizar alguma das técnicas de estimação descritas no capítulo anterior.

Trata-se exatamente da mesma situação descrita na Seção 4.2.1 e, portanto, poderíamos utilizar o mesmo exemplo: a aplicação de uma prova, composta de 30 itens novos (ou seja, com 30 itens que desejamos calibrar), aos alunos da 4.ª série diurna da rede pública estadual de São Paulo.

4.3.2 Quando todos os itens já estão calibrados

Este é o caso em que todos os itens já foram calibrados anteriormente, ou seja, quando não desejamos calibrar nenhum dos itens e estamos interessados apenas em estimar as habilidades dos respondentes. Este é um caso também bastante frequente na TRI, devido ao impulso que esta teoria deu na criação de *bancos de itens*. Tais bancos são formados por conjuntos de itens que já foram testados e calibrados a partir de um número significativo de indivíduos de uma dada população. Desta maneira, assumimos que os parâmetros desses itens já

são "conhecidos", ou seja, assumimos que conhecemos os verdadeiros valores dos parâmetros desses itens e assim, sempre que desejarmos, podemos aplicar novamente alguns desses itens do banco a outros indivíduos (ou até mesmo a um *único* indivíduo) e poderemos então estimar apenas suas habilidades, que estarão sempre na mesma métrica dos parâmetros dos itens.

A questão da métrica é um ponto que deve ser considerado com bastante cuidado numa situação como esta. Quando se "constrói"um banco de itens, uma informação fundamental é a *escala* em que aqueles itens foram calibrados. Isto porque as habilidades de indivíduos que serão estimadas futuramente a partir daqueles itens estarão nesta mesma métrica e portanto, quaisquer comparações diretas só poderão ser feitas com outro sujeitos que também tenham suas habilidades nesta escala.

Assim, para resolver este problema, basta utilizar um dos processos de estimação das habilidades dos indivíduos quando os parâmetros dos itens já são conhecidos, que foram descritos na Seção 3.3 do Capítulo 3.

Um exemplo para este tipo de situação seria a aplicação de uma prova, composta de 30 itens de 4.^{a} série que já foram calibrados numa aplicação anterior (por exemplo, numa aplicação de nível nacional como o SAEB), aos alunos da 4.^{a} série da rede pública estadual de São Paulo. Este tipo de procedimento é bastante comum, e nesse caso, o objetivo seria comparar a rede pública paulista com o desempenho nacional.

4.3.3 Quando alguns itens são novos e outros já estão calibrados

Neste caso, temos itens "novos"e itens já calibrados, ou seja, desejamos calibrar alguns itens e manter os parâmetros de outros, que já foram calibrados anteriormente. Este também é uma situação que está tipicamente ligada à criação de *bancos de itens*. Isto porque um banco de itens está continuamente em formação, ou seja, é bastante comum estarmos interessados em acrescentar novos itens ao conjunto que já se encontra no banco (assim como também é comum a retirada de itens do banco). Neste caso, o problema fundamental é garantir que os itens novos sejam calibrados na mesma métrica em que estão os outros itens do banco.

Na prática, este é um problema de solução mais complexa do que possa parecer em princípio. Isto porque é indispensável o uso de programas computacionais especificamente desenvolvidos para a análise de itens via TRI e esses

programas ainda apresentam algumas dificuldades com relação a situações como essa. Vamos comentar especificamente os problemas que podem surgir em casos como esse no Capítulo 7.

Um exemplo para esse caso seria a aplicação de uma prova, composta de 30 itens, aos alunos da 4.ª série diurna da rede pública estadual de São Paulo. Desses 30 itens, 15 são itens novos e 15 são itens que já foram calibrados numa aplicação de nível nacional do SAEB. Na prática, esta é uma situação bastante comum, pois quando são feitas avaliações regionais, por um lado há o interesse em criar e aplicar itens novos, mas por outro lado, há também o interesse em que os resultados obtidos possam ser comparados aos resultados nacionais.

Ilustramos até aqui, os casos (a), (b) e (c) considerando-se a situação 1. As outras situações onde tratamos apenas de uma população (situações 2 e 3), são análogas. No entanto, quando temos duas (ou mais) populações envolvidas (situações 4 e 6), e desejamos estimar itens novos e manter fixos os parâmetros dos itens já calibrados (caso (c)), poderemos ter problemas com a métrica. Os casos (a) e (b) não apresentam problemas, sendo análogos à situação anterior.

Sempre que há mais de uma população envolvida nos processos de estimação, como já foi comentado anteriormente, existem problemas de indeterminação de escala. Para resolver este problema, devemos definir uma das populações como sendo a referência, e então, as demais populações serão posicionadas com relação a ela.

Este tipo de problema sempre irá ocorrer quando fazemos a equalização entre duas ou mais populações *durante* o processo de estimação dos itens. Uma outra maneira de solucionarmos o problema seria através da chamada equalização a posteriori, que será discutida a seguir.

4.4 Equalização a posteriori

Até aqui discutimos formas de equalização entre 2 ou mais populações feitas *durante* o próprio processo de estimação dos parâmetros. Mas, também é possível fazer a equalização *a posteriori*, isto é, depois de terminado o processo de calibração dos itens. Basicamente, a equalização a posteriori é feita da seguinte maneira: calibra-se separadamente os dois conjuntos de itens, que foram submetidos às duas populações de interesse. Obviamente, a condição necessária é que hajam itens comuns entre os dois conjuntos. Assim, para os

itens comuns, teremos dois conjuntos de estimativas, cada uma na métrica de suas respectivas populações. Daí, através dessas duas estimativas para os itens comuns estabelece-se algum tipo de *relação* que permita colocarmos os parâmetros de um dos conjuntos de itens na escala do outro. Com todos os itens na mesma métrica, pode-se então estimar as habilidades de todos os respondentes, que então estarão também na mesma escala.

Pela propriedade de invariância, já discutida no Capítulo 2, dado que o modelo é adequado aos dados, os parâmetros a e b de um certo item apresentado a 2 grupos de respondentes devem satisfazer, a menos de flutuações amostrais, as seguintes relações lineares:

$$b_{G1} = \alpha b_{G2} + \beta \quad \text{e} \quad a_{G1} = \frac{1}{\alpha} a_{G2}, \tag{4.1}$$

onde b_{G1} e b_{G2} são os valores do parâmetro de dificuldade e a_{G1} e a_{G2} são os valores do parâmetro de discriminação nos grupos 1 e 2, respectivamente. Uma vez determinados os coeficientes α e β , as estimativas dos parâmetros dos itens do grupo 2 podem facilmente ser colocados na mesma escala das estimativas do grupo 1.

Vários métodos, que se baseiam nessas relações lineares existentes entre os parâmetros de um mesmo item medidos em escalas diferentes, poderiam ser então utilizados para determinar os coeficientes α e β. A solução mais natural — pelo próprio tipo de relação existente entre os parâmetros — seria determinar esses coeficientes através de uma regressão linear simples. No entanto, a crítica feita à utilização desse método é que ele *não é simétrico*, ou seja, uma regressão de x por y é diferente de uma regressão de y por x.

Um dos métodos de equalização a posteriori existentes que não apresenta esse problema, ou seja, é invariante (simétrico) em relação às variáveis utilizadas, é denominado *Média-Desvio (Mean-Sigma)*. O método Média-Desvio utiliza:

$$\alpha = \frac{S_{G1}}{S_{G2}} \quad \text{e} \quad \beta = M_{G1} - \alpha M_{G2}, \tag{4.2}$$

onde S_{G1} e S_{G2} são os desvios-padrão e M_{G1} e M_{G2} as médias amostrais das estimativas dos parâmetros de dificuldade dos itens comuns nos grupos 1 e 2, respectivamente. Da mesma forma, as habilidades dos respondentes do grupo

2 podem ser colocadas na mesma escala das habilidades dos respondentes do grupo 1 a partir da relação

$$\theta^1_{G2} = \alpha\theta_{G2} + \beta, \tag{4.3}$$

onde θ^1_{G2} é o valor da habilidade θ_{G2} na escala do grupo 1. Maiores detalhes sobre este e outros métodos de equalização, como por exemplo *Média-Desvio Robusto* e *Curva Característica*, podem ser encontrados em Kolen & Brennan (1995).

Exemplificando, uma avaliação feita no estado do Rio Grande do Norte (ver Fundação Carlos Chagas (1997)) utilizou alguns itens do SAEB 95, com o intuito de colocar os resultados obtidos na mesma métrica do SAEB. As Fuguras 4.2 e 4.3 mostram as relações entre as estimativas dos parâmetros a e b nas duas avaliações, para a disciplina Língua Portuguesa da 8.^{a} série do Ensino Fundamental.

Figura 4.2 *Gráfico de dispersão das estimativas do parâmetro de dificuldade - b dos itens comuns da prova de Língua Portuguesa da* 8.^{a} *série entre o RN e o SAEB*

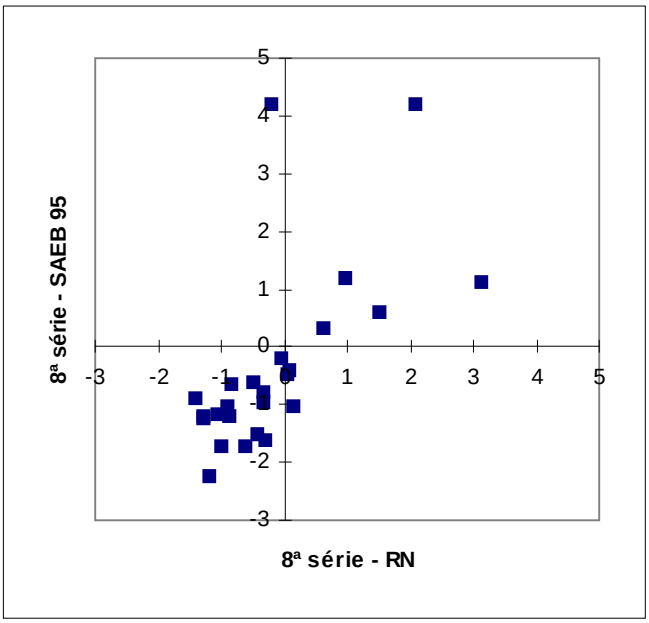
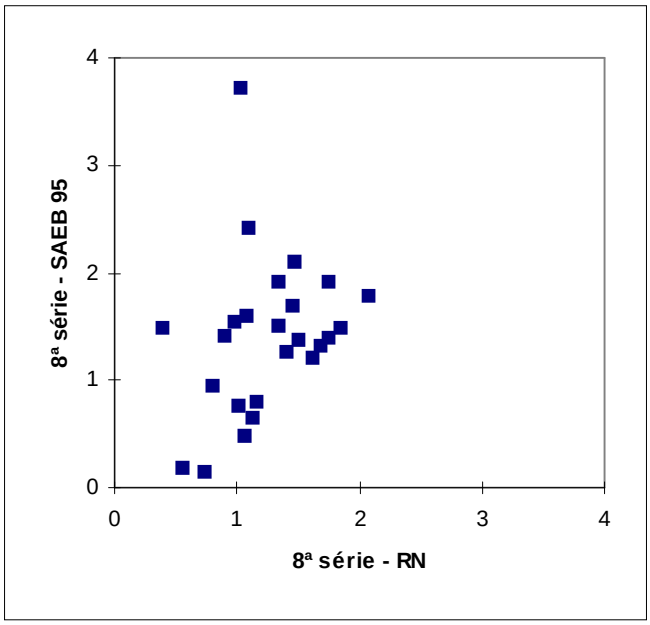


Figura 4.3 *Gráfico de dispersão das estimativas do parâmetro de discriminação - a dos itens comuns da prova de Língua Portuguesa da 8.ª série entre o RN e o SAEB*



Utilizando o método Média-Desvio, os coeficientes α e β obtidos foram:

$$\alpha = \frac{S_{SAEB}}{S_{RN}} = \frac{1,614}{1,104} = 1,462,$$

$$\beta = M_{SAEB} - \alpha M_{RN} = -0,363 - 1,462 \times -0,162 = -0,126.$$

Logo, as estimativas dos parâmetros obtidas na avaliação feita com os alunos do Rio Grande do Norte foram colocadas na mesma métrica do SAEB 95 através das seguintes expressões:

$$a_{RN}^{NOVO} = \frac{1}{\alpha} a_{RN} = \frac{1}{1,462} a_{RN},$$

$$b_{RN}^{NOVO} = \alpha b_{RN} + \beta = 1,462 b_{RN} - 0,126,$$

$$\theta_{RN}^{NOVO} = \alpha\theta_{RN} + \beta = 1,462\theta_{RN} - 0,126.$$

Uma última observação sobre equalização deve ser feita com relação à *quantidade* de itens comuns. Certamente, quanto maior o número de itens comuns, melhor será a qualidade da equalização. Assim, o melhor caso de equalização entre dois grupos distintos é a situação da Seção 4.2.4, ou seja, quando trata-se exatamente da mesma prova. No entanto, já sabemos que não é necessário que todos os itens sejam comuns. O número mínimo de itens comuns necessário para uma boa equalização entre duas populações depende basicamente de dois fatores: do tipo de equalização que será feita e da "qualidade"desses itens comuns.

Equalizações feitas durante o processo de calibração, com os modelos para duas ou mais populações que serão discutidos no próximo capítulo, são mais "eficazes"e portanto, exigem um número menor de itens comuns do que equalizações feitas a posteriori. Além disso, se os itens comuns utilizados na equalização tiverem níveis de dificuldade baixos ou altos demais com relação às populações envolvidas, ou então se apresentarem baixo poder de discriminação, haverá necessidade de um número maior de itens.

Alguns autores têm sugerido pelo menos 6 itens comuns entre 2 provas de 30 itens, quando a equalização é feita durante a calibração. Um estudo de simulação considerando diferentes situações de equalização pode ser encontrado em Andrade (1999).

Capítulo **5**

Estimação: duas ou mais populações

5.1 Introdução

Como descrito no capítulo anterior, é freqüente a situação em que temos duas ou mais populações envolvidas na análise. Estas populações podem ser caracterizadas por diferentes graus de escolaridade, região, sexo, tipo de escola, etc. O primeiro passo para que os resultados relativos às várias populações possam ser comparáveis é a exigência de itens comuns nos testes aplicados a estas populações, criando uma estrutura de ligação entre as mesmas. Nessa situação, o procedimento usual é fazer a estimação para cada população e utilizar uma das técnicas de equalização descritas na Seção 4.3.

Um abordagem alternativa é o Modelo para Várias Populações proposto por Bock & Zimowski (1997), introduzido na Secção 2.3, que representou um grande avanço na TRI. Nesse modelo considera-se que há K populações independentes em estudo e é feita uma análise conjunta das respostas amostrais dessas populações. Considera-se que a distribuição da habilidade dos indivíduos da população k segue uma determinada distribuição com vetor de parâmetros $\boldsymbol{\eta}_k$. Frequentemente adota-se a distribuição Normal com $\boldsymbol{\eta}_k = (\mu_k, \sigma_k^2)'$, sendo que estes parâmetros representam, respectivamente, a média e a variância das habilidades da população k, $k = 1, \cdots, K$.

A grande vantagem da abordagem de Bock & Zimowski está no fato que a equalização é feita automaticamente no próprio processo de estimação. Desta forma, não estamos mais sujeitos a diferenças nas estimativas dos parâmetros devidas ao método de equalização escolhido. Além disso, na presença de várias populações (digamos, $K \geq 5$), com as equalizações sendo feitas entre os testes k e $k+1$, temos erros (relativos à regressão, por exemplo) associados a cada equalização entre duas populações, que serão acumulados para a estimação de (μ_2,σ_2^2), (μ_3,σ_3^2), $\cdots$, e principalmente de (μ_K,σ_K^2), podendo levar a uma má

estimação destes parâmetro. Além disso, essa abordagem requer um número menor de itens comuns, em comparação com outros métodos, para produzir resultados similares, conforme discutido no capítulo anterior.

Sejam u_{kji} a resposta (binária) ao item i oriunda do j-ésimo indivíduo do grupo k, e θ_{kj} a habilidade do j-ésimo indivíduo do grupo k. (Por *grupo* k entenderemos a amostra relativa à população k.) Embora no desenvolvimento que segue a função de resposta possa assumir qualquer uma das formas descritas no Capítulo 2, para fins de aplicação utilizaremos a função ML3, que tem sido a função mais utilizada pelos pesquisadores da área, dada abaixo

$$P(u_{kji} = 1|\theta_{kj}) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_{kj} - b_i)}}.$$

Algumas suposições serão necessárias para a construção do modelo. Além da independência local, assumiremos que as respostas oriundas de indivíduos diferentes serão independentes. Vamos considerar a mesma função de resposta para todos os itens.

5.2 Notações e definições

Embora tenhamos K testes, devemos notar que alguns itens estarão em dois ou mais testes. Por conta disso vamos fazer uma ordenação nos I itens que compõem o conjunto dos K testes, representando-os por $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_I)$ e denotando por $\boldsymbol{\mathcal{I}}_k$ o conjunto dos índices dos itens aplicados ao grupo k. Considerando I_k o número de itens no teste k, teremos que $I \leq \sum_{k=1}^{K} I_k$. Sejam n_k o número de indivíduos do grupo k e n o número total de indivíduos na amostra. Sejam ainda, $\boldsymbol{U}_{kj.} = (U_{kj1}, U_{kj2}, \cdots, U_{kjI_k})$ o vetor aleatório de respostas do indivíduo j do grupo k; $\boldsymbol{U}_{k..} = (\boldsymbol{U}_{1..}, \boldsymbol{U}_{2..}, \cdots, \boldsymbol{U}_{n_k..})$ o vetor aleatório de respostas do grupo k e $\boldsymbol{U}_{...} = (\boldsymbol{U}_{1..}, \boldsymbol{U}_{2..}, \cdots, \boldsymbol{U}_{n..})$ o vetor total de respostas. De forma similar, representaremos as respostas observadas por $u_{kji}, \boldsymbol{u}_{kj.}, \boldsymbol{u}_{k..}$ e $\boldsymbol{u}_{...}$. Com esta notação e a independência local, podemos escrever a probabilidade associada ao vetor de respostas $\boldsymbol{U}_{kj.}$ como

$$P(\boldsymbol{u}_{kj.}|\theta_{kj}, \boldsymbol{\zeta}) = \prod_{i \in \boldsymbol{\mathcal{I}}_k} P(u_{kji}|\theta_{kj}, \boldsymbol{\zeta}_i).$$

Como comentado no Capítulo 3, o método da Máxima Verossimilhança Marginal, bem como o Bayesiano, têm sido preferidos ao método da Máxima Verossimilhança Conjunta para a estimação dos parâmetros de interesse. Além disso, o fato de podermos associar distribuições para a habilidade da população em estudo nos permite criar estruturas para os parâmetros das respectivas funções densidade de probabilidade, que serão fundamentais nesse modelo. De forma geral, consideremos que as habilidades dos indivíduos da população k, θ_{jk}, $j = 1, \cdots, n_k$, são realizações de uma variável aleatória, θ_k, com distribuição contínua e função densidade de probabilidade $g(\theta|\boldsymbol{\eta}_k)$, duplamente diferenciável, com as componentes de $\boldsymbol{\eta}_k$ conhecidas e finitas. Para o caso em que θ_k tem distribuição Normal, temos $\boldsymbol{\eta}_k = (\mu_k, \sigma_k^2)$, onde μ_k é a média e σ_k^2 a variância das habilidades dos indivíduos da população k, $k = 1, \cdots, K$.

Na situação em que temos uma única população em estudo, não há necessidade de estimação dos parâmetros populacionais. Isso ocorre porque a métrica é estabelecida fixando-se os parâmetros populacionais, geralmente em $\mu = 0$ e $\sigma = 1$, onde μ é a média e σ é o desvio-padrão das habilidade da população considerada. Na presença de várias populações, temos mais um conjunto de parâmetro a estimar: $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_K)$, que serão referidos como *Parâmetros Populacionais*. Entretanto, ainda há a necessidade do estabelecimento da métrica e isso pode ser resolvido fixando-se os parâmetros relativos a qualquer uma das populações. Neste livro adotaremos a seguinte referência:

$$\mu_1 = 0, \qquad \sigma_1 = 1. \tag{5.1}$$

Logo, resta apenas a estimação de $\boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_K$. Novamente, a estimação neste modelo é feita por máxima verossimilhança marginal, com o diferencial que a primeira etapa envolve a estimação dos parâmetros dos itens e dos parâmetros populacionais; as habilidades individuais são estimadas na segunda etapa. Cabe notar aqui uma grande contribuição do modelo de Bock & Zimowski, a de que as médias populacionais podem ser estimadas diretamente, ao passo que o procedimento anterior era fazer a estimação das habilidades para cada grupo, adotar um método de equalização para colocá-las na mesma escala e, finalmente, obter a média amostral das habilidades de cada grupo.

Como faremos a estimação por máxima verossimilhança marginal, haverá alguma similaridade com o desenvolvimento da Seção 3.5. Porém, devido a

importância deste modelo, a maioria dos detalhes serão apresentados. Para ressaltar a diferença nos desenvolvimentos das equações de estimação para os parâmetros dos itens e populacionais, abordaremos a estimação dos parâmetros dos itens na Seção 5.3 e dos parâmetros populacionais na Seção 5.4. As equações para a estimação conjunta dos parâmetros dos itens e populacionais será o conjunto das equações desenvolvidas nas duas referidas seções.

5.3 Estimação dos parâmetros dos itens

Embora as equações de estimação a serem desenvolvidas nesta seção tenham como prioridades compor o conjunto das equações para estimação conjunta dos parâmetros dos itens e populacionais, cabe notar que elas também poderão ser adotadas na situação em que os parâmetros populacionais são conhecidos. Uma situação dessas ocorre quando tais parâmetros populacionais foram estimados em outra análise, talvez com um número de indivíduos bem maior, de forma que não há interesse na reestimação desses parâmetros.

Com as notações definidas acima, temos que a probabilidade marginal de $\boldsymbol{U}_{kj.}$ é dada por

$$\begin{aligned} P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta},\boldsymbol{\eta}_k) &= \int_{I\!R} P(\boldsymbol{u}_{kj.}|\theta,\boldsymbol{\zeta},\boldsymbol{\eta}_k) g(\theta|\boldsymbol{\eta}_k) d\theta \\ &= \int_{I\!R} P(\boldsymbol{u}_{kj.}|\theta,\boldsymbol{\zeta}) g(\theta|\boldsymbol{\eta}_k) d\theta, \end{aligned}$$

onde na última igualdade usamos que a distribuição de $\boldsymbol{U}_{kj.}$ não é função de $\boldsymbol{\eta}_k$.

Usando a independência entre as respostas de diferentes indivíduos, podemos escrever a probabilidade associada ao vetor de respostas $\boldsymbol{U}_{...}$ como

$$P(\boldsymbol{u}_{...}|\boldsymbol{\zeta},\boldsymbol{\eta}) = \prod_{k=1}^{K} \prod_{j=1}^{n_k} P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta},\boldsymbol{\eta}_k). \tag{5.2}$$

Embora a verossimilhança possa ser escrita como (5.2), tem sido freqüente utilizar a abordagem de *Padrões de Resposta*. Como temos I_k itens no teste

k, com duas possíveis respostas para cada item (0 ou 1), há $S_k = 2^{I_k}$ possíveis respostas (padrões de resposta) associados a esse teste. Seja r_{kj} o número de ocorrências distintas do padrão de resposta j no grupo k, e ainda $s_k \leq \min(n_k, S_k)$ o número de padrão de respostas com $r_{kj} > 0$. Segue que

$$\sum_{j=1}^{s_k} r_{kj} = n_k. \tag{5.3}$$

Pela independência entre as respostas dos diferentes indivíduos, temos que os dados seguem uma distribuição $Produto - Multinomial$, isto é,

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \prod_{k=1}^{K} \left\{ \frac{n_k!}{\prod_{j=1}^{s_k} r_{jk}!} \prod_{j=1}^{s_k} [P(\boldsymbol{u}_{jk.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)]^{r_{jk}} \right\}. \tag{5.4}$$

E, portanto, a log-verossimilhança é

$$\log L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{k=1}^{K} \log \left\{ \frac{n_k!}{\prod_{j=1}^{s_k} r_{jk}!} \right\} + \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{jk} \log P(\boldsymbol{u}_{jk.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k). \tag{5.5}$$

As equações de estimação para os parâmetros dos itens são dadas por

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \mathbf{0}, \quad i = 1, \cdots, I, \tag{5.6}$$

com

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} &= \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left\{ \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{jk} \log P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k) \right\} \\ &= \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{jk} \frac{1}{P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} \frac{\partial P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)}{\partial \boldsymbol{\zeta}_i} \\ &= \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_{kj}^*(\theta) d\theta, \end{aligned} \tag{5.7}$$

onde

$$g^*_{kj}(\theta) \equiv g(\theta|\boldsymbol{u}_{kj.}, \boldsymbol{\zeta}, \boldsymbol{\eta}_k) = \frac{P(\boldsymbol{u}_{kj.}|\theta, \boldsymbol{\zeta}) g(\theta|\boldsymbol{\eta}_k)}{P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)}. \tag{5.8}$$

As equações específicas para cada parâmetro do vetor $\boldsymbol{\zeta}_i = (a_i, b_i, c_i)'$ podem então ser obtidas de (5.7). Para o parâmetro de discriminação a_i, usando também (3.8), obtem-se

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial a_i} &= \\
&= \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) \left(\frac{\partial P_i}{\partial a_i} \right) \frac{W_i}{P^*_i Q^*_i} \right] g^*_{kj}(\theta) d\theta \\
&= \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) D(1 - c_i)(\theta - b_i) P^*_i Q^*_i \frac{W_i}{P^*_i Q^*_i} \right] g^*_{kj}(\theta) d\theta \\
&= D(1 - c_i) \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i)(\theta - b_i) W_i \right] g^*_{kj}(\theta) d\theta.
\end{aligned}$$

Para o parâmetro de dificuldade b_i, usando também (3.9), obtem-se

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial b_i} &= \\
&= \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) \left(\frac{\partial P_i}{\partial b_i} \right) \frac{W_i}{P^*_i Q^*_i} \right] g^*_{kj}(\theta) d\theta \\
&= \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i)(-1) D a_i (1 - c_i) P^*_i Q^*_i \frac{W_i}{P^*_i Q^*_i} \right] g^*_{kj}(\theta) d\theta \\
&= -D a_i (1 - c_i) \sum_{k=1}^{K} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) W_i \right] g^*_{kj}(\theta) d\theta.
\end{aligned}$$

Por último, para o parâmetro de acerto ao acaso c_i, usando também (3.10), obtem-se

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial c_i} &= \sum_{k=1}^{K}\sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) \left(\frac{\partial P_i}{\partial c_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_{kj}^*(\theta) d\theta \\
&= \sum_{k=1}^{K}\sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) Q_i^* \frac{W_i}{P_i^* Q_i^*} \right] g_{kj}^*(\theta) d\theta \\
&= \sum_{k=1}^{K}\sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) \frac{W_i}{P_i^*} \right] g_{kj}^*(\theta) d\theta.
\end{aligned}$$

Em resumo, as equações de estimação para a_i, b_i e c_i são, respectivamente,

$$a_i \;:\; D(1-c_i) \sum_{k=1}^{K}\sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i)(\theta - b_i) W_i \right] g_{kj}^*(\theta) d\theta = 0, \quad (5.9)$$

$$b_i \;:\; -D a_i (1-c_i) \sum_{k=1}^{K}\sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) W_i \right] g_{kj}^*(\theta) d\theta = 0, \quad (5.10)$$

$$c_i \;:\; \sum_{k=1}^{K}\sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[(u_{kji} - P_i) \frac{W_i}{P_i^*} \right] g_{kj}^*(\theta) d\theta = 0, \quad (5.11)$$

as quais não possuem solução explícita.

5.4 Estimação dos parâmetros populacionais

Novamente, embora as equações de estimação a serem desenvolvidas nesta seção tenham como prioridades compor o conjunto das equações para estimação conjunta dos parâmetros dos itens e populacionais, cabe notar que elas também poderão ser adotadas na situação em que os parâmetros dos itens são conhecidos.

Considerando a log-verossimilhança obtida em (5.5), as equações de estimação para as habilidades médias e variâncias das populações são obtidas por

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \mu_k} = 0 \qquad \text{e} \qquad \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \sigma_k^2} = 0, \quad k = 2, \cdots, K.$$

Mas,

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}_k} &= \sum_{j=1}^{s_k} r_{jk} \frac{1}{P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} \int_{I\!R} P(\boldsymbol{u}_{kj.}|\theta, \boldsymbol{\zeta}) \left(\frac{\partial g(\theta|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right) d\theta \\
&= \sum_{j=1}^{s_k} r_{jk} \frac{1}{P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} \int_{I\!R} P(\boldsymbol{u}_{kj.}|\theta, \boldsymbol{\zeta}) \left(\frac{\partial \log g(\theta|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right) g(\theta|\boldsymbol{\eta}_k) d\theta \\
&= \sum_{j=1}^{s_k} r_{jk} \int_{I\!R} \left(\frac{\partial \log g(\theta|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right) g_{kj}^*(\theta) d\theta.
\end{aligned}$$

Se utilizarmos a distribuição $N(\mu_k, \sigma_k^2)$ para θ_k, teremos que

$$\frac{\partial \log g(\theta|\boldsymbol{\eta}_k)}{\partial \mu_k} = \frac{\theta - \mu_k}{\sigma_k^2}$$

e

$$\frac{\partial \log g(\theta|\boldsymbol{\eta}_k)}{\partial \sigma_k^2} = -\frac{\sigma_k^2 - (\theta - \mu_k)^2}{2\sigma_k^4}.$$

Assim, as formas finais das equações de estimação são

$$\mu_k \quad : \qquad (\sigma_k^2)^{-1} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} (\theta - \mu_k) g_{kj}^*(\theta) d\theta = 0, \tag{5.12}$$

$$\sigma_k^2 \quad : \qquad -(2\sigma_k^4)^{-1} \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \left[\sigma_k^2 - (\theta - \mu_k)^2 \right] g_{kj}^*(\theta) d\theta = 0. \tag{5.13}$$

Note que, se fizermos

$$\mu_{kj} = \int_{I\!R} \theta g_{kj}^*(\theta) d\theta, \tag{5.14}$$

$$\sigma_{kj}^2 = \int_{I\!R} (\theta - \mu_{kj})^2 g_{kj}^*(\theta) d\theta, \tag{5.15}$$

que representam a média e a variância da distribuição condicional da habilidade da população k, dado $\{\boldsymbol{U}_{kj.} = \boldsymbol{u}_{kj.}\}$, respectivamente, então, por (5.3), (5.12) e (5.14), segue que

$$\begin{aligned} 0 &= \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} \theta g_{kj}^*(\theta) d\theta - \widehat{\mu}_k \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} g_{kj}^*(\theta) d\theta \\ &= \sum_{j=1}^{s_k} r_{kj} \widehat{\mu}_{kj} - \widehat{\mu}_k \sum_{j=1}^{s_k} r_{kj} \\ &= \sum_{j=1}^{s_k} r_{kj} \widehat{\mu}_{kj} - n_k \widehat{\mu}_k, \end{aligned}$$

de onde concluímos que

$$\widehat{\mu}_k = \frac{1}{n_k} \sum_{j=1}^{s_k} r_{kj} \widehat{\mu}_{kj}. \tag{5.16}$$

Também, por (5.13), (5.15) e usando que $\theta - \mu_k = (\theta - \mu_{kj}) + (\mu_{kj} - \mu_k)$, temos

$$\begin{aligned} 0 &= \sum_{j=1}^{s_k} r_{kj} \widehat{\sigma}_k^2 \int_{I\!R} g_{kj}^*(\theta) d\theta - \sum_{j=1}^{s_k} r_{kj} \int_{I\!R} (\theta - \widehat{\mu}_k)^2 g_{kj}^*(\theta) d\theta \\ &= \widehat{\sigma}_k^2 \sum_{j=1}^{s_k} r_{kj} - \sum_{j=1}^{s_k} r_{kj} \left[\int_{I\!R} (\theta - \widehat{\mu}_{kj})^2 g_{kj}^* d\theta + \int_{I\!R} (\widehat{\mu}_{kj} - \widehat{\mu}_k)^2 g_{kj}^*(\theta) d\theta \right] \\ &= n_k \widehat{\sigma}_k^2 - \sum_{j=1}^{s_k} r_{kj} \left[\widehat{\sigma}_{kj}^2 + (\widehat{\mu}_{kj} - \widehat{\mu}_k)^2 \right], \end{aligned}$$

de onde concluímos que

$$\widehat{\sigma}_k^2 = \frac{1}{n_k} \sum_{j=1}^{s_k} r_{kj} \left[\widehat{\sigma}_{kj}^2 + (\widehat{\mu}_{kj} - \widehat{\mu}_k)^2 \right]. \tag{5.17}$$

Note que $g^*_{kj}(\theta)$ depende dos parâmetros dos itens e também dos parâmetros populacionais e, consequentemente, seu valor nas expressões acima deve ser calculado a partir de estimativas desses parâmetros.

Representando por $\overline{\mu}_{k.}$ a média das esperanças condicionais μ_{kj}, por $\overline{\sigma}^2_{k.}$ a média das variâncias condicionais σ^2_{kj} e por $\delta^2_{k.}$ uma medida adequada de variabilidade entre as médias condicionais, todas associadas ao grupo k, ou seja,

$$\overline{\mu}_{k.} = \frac{1}{n_k}\sum_{j=1}^{s_k} r_{kj}\mu_{kj}, \quad \overline{\sigma}^2_{k.} = \frac{1}{n_k}\sum_{j=1}^{s_k} r_{kj}\sigma^2_{kj} \quad \text{e} \quad \delta^2_{k.} = \frac{1}{n_k}\sum_{j=1}^{s_k} r_{kj}(\mu_{kj}-\mu_k)^2,$$

podemos escrever as equações (5.16) e (5.17) como

$$\widehat{\mu}_k = \widehat{\overline{\mu}}_{k.} \quad \text{e} \quad \widehat{\sigma}^2_k = \widehat{\overline{\sigma}}^2_{k.} + \widehat{\delta_{k.}}^2, \quad k = 2, \cdots, K. \tag{5.18}$$

Estas expressões nos permitem interpretações bastante intuitivas. Primeiro, notemos que os somatórios nas definições acima podem ser adaptados de forma a considerar as respostas individuais ao invés dos padrões de respostas. Com isso, o estimador para a habilidade média da população k é a média obtida com os estimadores das médias da distribuição condicional da habilidade, dados os vetores de respostas individuais $\boldsymbol{u}_{kj.}$. Por outro lado, o estimador para a variância das habilidades da população k não é simplesmente a média entre estimadores das variâncias da distribuição condicional da habilidade, dados os vetores de respostas individuais $\boldsymbol{u}_{kj.}$. Existe também uma outra contribuição relativa à variabilidade entre os estimadores das médias da distribuição condicional da habilidade com relação ao estimador da média populacional associada.

5.4.1 Estimação conjunta: aplicação do algoritmo EM

De forma similar ao que foi feito na Seção 3.5.5, podemos escrever a equação de estimação para $\boldsymbol{\zeta}_i$ como

$$\boldsymbol{\zeta}_i \quad : \quad \sum_{k=1}^{K}\int_{I\!R}\left[(r_{ki}(\theta) - P_i f_{ki}(\theta))\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)\frac{W_i}{P^*_i Q^*_i}\right] d\theta = \mathbf{0}, \tag{5.19}$$

onde

$$f_{ki}(\theta) = \sum_{j=1}^{s_k} r_{kj} g_j^*(\theta) e \quad r_{ki}(\theta) = \sum_{j=1}^{s_k} r_{kj} u_{kji} g_{kj}^*(\theta)$$

representam, respectivamente, o números de indivíduos do grupo k com habilidade θ respondendo ao item i e o números destes indivíduos que respondem corretamente ao item i. Novamente, as integrais são aproximadas através de quadratura Gaussiana. Fixados os q nós $\overline{\theta}_{kl}$ e os pesos $A_{kl},\ l = 1, \cdots, q,\ k = 1, \cdots, K$, e com estimativas iniciais dos parâmetros dos itens, $\widehat{\boldsymbol{\zeta}}_i,\ i = 1, \cdots, I$, as equações (5.18) podem ser resolvidas diretamente para obtenção das estimativas desejadas. A estimação é feita em separado para cada item, e por isso poderemos utilizar o desenvolvimento da Seção 3.2. Reformulando-se os passos do algoritmo EM descritos na Seção 3.5.3, para a situação de duas ou mais populações, teremos

Passo E

1. Usar os pontos de quadratura $\overline{\theta}_{kl}$, os pesos $A_{kl},\ l = 1, \cdots, q$ e estimativas iniciais dos parâmetros dos itens , $\boldsymbol{\zeta}_i,\ i = 1, \cdots, I$, e dos parâmetros populacionais, $\boldsymbol{\mu}_k$ e $\sigma_k^2,\ k = 1, \cdots, K$, para gerar $g_{kj}^*(\overline{\theta}_{kl})$ e, posteriormente, $\overline{r}_{kli}$ e $\overline{f}_{kli}$, $i = 1, \cdots, I$ e $k = 1, \cdots, q$.
2. Usar os pontos de quadratura e $g_{kj}^*(\overline{\theta}_{kl})$ para obter $\widehat{\mu}_{kj}$ e $\widehat{\sigma}_{kj}^2$ por (5.14) e (5.15), e poteriormente, $\widehat{\mu}_k$ e $\widehat{\sigma}_k^2$ por (5.18).

Passo M Com $\boldsymbol{r}$, $\boldsymbol{f}$ e $\boldsymbol{\eta}$ obtidos no Passo E, resolver as equações de estimação para $\boldsymbol{\zeta}_i,\ i = 1, \cdots, I$, usando o algoritmo Newton-Raphson ou "Scoring"de Fisher através das expressões da Seção 3.2.

Estes passos compõem cada iteração do algoritmo EM, as quais serão repetidas até que algum critério de parada seja alcançado. Após a finalização do processo, os erros-padrão são obtidos com o uso de (3.29).

Devemos notar que no passo M as expressões para a maximização são um pouco modificadas, com relação às expressões da Seção 3.2, devido a introdução de novos grupos. Se $\widehat{\boldsymbol{\zeta}}_i^{(t)}$ é uma estimativa de $\boldsymbol{\zeta}_i$ na iteração t, o processo iterativo de Newton-Raphson para obtenção de $\widehat{\boldsymbol{\zeta}}_i^{(t+1)}$ é dado pela expressão (3.25), onde

$$\begin{aligned}
\boldsymbol{h}(\boldsymbol{\zeta}_i) &= \sum_{k=1}^{K}\sum_{l=1}^{q}(r_{kli} - f_{kli}P_{kli})W_{kli}\boldsymbol{h}_{kli}, \\
\boldsymbol{H}(\boldsymbol{\zeta}_i) &= \sum_{k=1}^{K}\sum_{l=1}^{q}(r_{kli} - f_{kli}P_{kli})W_{kli}\left\{\boldsymbol{H}_{kli} - (r_{kli} - f_{kli}P_{kli})W_{kli}\boldsymbol{h}_{kli}\boldsymbol{h}'_{kli}\right\},
\end{aligned}$$

com $P_{kli}, W_{kli}, \boldsymbol{H}_{kli}$ e $\boldsymbol{h}_{kli}$ similares à Seção 3.2, com θ_k substituída por $\overline{\theta}_{kl}$. Para a aplicação do método "Scoring"de Fisher, devemos substituir $\boldsymbol{H}(\boldsymbol{\zeta}_i)$ pelo seu valor esperado, ou seja,

$$\boldsymbol{\Delta}(\boldsymbol{\zeta}_i) = -\sum_{k=1}^{K}\sum_{l=1}^{q}\left\{P^*_{kli}Q^*_{kli}W_{kli}\boldsymbol{h}_{kli}\boldsymbol{h}'_{kli}\right\}.$$

5.5 Estimação bayesiana dos parâmetros dos itens

Como podemos perceber no caso da estimação por MVM, as equações de estimação quando temos várias populações em estudo são bastante similares ao caso em que temos apenas uma população em estudo, com algumas modificações das componentes devidas à presença de outras populações. Isso também se reflete no caso da estimação bayesiana, onde são utilizadas as equações de estimação obtidas por MVM, com o incremento de parcelas relativas às distribuições a priori adotadas para os parâmetros dos itens. Neste capítulo, vamos adotar as mesmas distribuições a priori consideradas na Seção 3.6.1.

Com isso, as equações de estimação podem ser escritas como

$$a_i: \quad D(1-c_i)\sum_{k=1}^{K}\int_{I\!R}(\theta-b_i)\left[r_{ki}(\theta)-P_i f_{ki}(\theta)\right]W_i d\theta - \frac{1}{a_i}\left[1+\frac{\log a_i-\mu_a}{\sigma_a^2}\right]=0,$$

$$b_i: \quad -Da_i(1-c_i)\sum_{k=1}^{K}\int_{I\!R}\left[r_{ki}(\theta)-P_i f_{ki}(\theta)\right]W_i d\theta - \frac{(b_i-\mu_b)}{\sigma_b^2}=0,$$

$$c_i: \quad \sum_{k=1}^{K}\int_{I\!R}\left[r_{ki}(\theta)-P_i f_{ki}(\theta)\right]\frac{W_i}{P_i^*}d\theta + \frac{\alpha-2}{c_i} - \frac{\beta-2}{1-c_i}=0.$$

As derivadas segundas destas expressões são facilmente obtidas pela Seção 3.2 e por (3.101). A aplicação do algoritmo EM se dá de forma idêntica à estimação por MVM, delineada na seção anterior.

5.6 Estimação das habilidades

Uma etapa que pode ser considerada opcional é a estimação das habilidades. Talvez o interesse da análise se concentre apenas na estimação dos parâmetros dos itens e populacionais, sem relevar a estimação das habilidades. Em caso contrário, as habilidades podem ser estimadas por MV, como descrito na Seção 3.3, ou de forma bayesiana, como descrito na Seção 3.6.2. Devido a presença de vários grupos na análise, as expressões para obtenção das habilidades são ligeiramente modificadas, e por isso as apresentaremos nesta seção.

Vale ressaltar que em todos os métodos de estimação descritos abaixo, consideraremos fixos os parâmetros dos itens e os populacionais.

5.6.1 Estimação por MV

Neste caso, a estimação das habilidades é feita iterativamente pelo algoritmo Newton-Raphson ou método "Scoring"de Fisher. Considerando $\widehat{\theta}_{kj}^{(t)}$ uma estimativa de θ_{kj} na iteração t, então na iteração $t+1$ teremos que

$$\widehat{\theta}_{kj}^{(t+1)} = \widehat{\theta}_{kj}^{(t)} - [H(\widehat{\theta}_{kj}^{(t)})]^{-1} h(\widehat{\theta}_{kj}^{(t)}), \tag{5.20}$$

onde

$$H(\theta_{kj}) \;=\; \sum_{i\in\boldsymbol{\mathcal{I}}_k}(u_{kji}-P_{kji})W_{kji}\left\{H_{kji}-(u_{kji}-P_{kji})W_{kji}h_{kji}^2\right\},$$

com $P_{kji}, W_{kji}, H_{kji}$ e h_{kji} similares à Seção 3.3, com θ_j substituída por θ_{kj}. Para aplicação do método "Scoring"de Fisher, devemos substituir $H(\theta_{kj})$ pelo seu valor esperado, ou seja,

$$\Delta(\theta_{kj}) \;=\; -\sum_{i\in\boldsymbol{\mathcal{I}}_k}P_{kji}^*Q_{kji}^*W_{kji}h_{kji}^2.$$

5.6.2 Estimação por MAP

A estimação pelo máximo da posteriori (MAP) também é obtida iterativamente pela aplicação de (5.20), com

$$H(\theta_{kj})=\sum_{i\in\boldsymbol{\mathcal{I}}_k}(u_{kji}-P_{kji})W_{kji}\left\{H_{kji}-(u_{kji}-P_{kji})W_{kji}h_{kji}^2\right\}-\frac{1}{\sigma_k^2}, \quad (5.21)$$

onde h_{kji} e H_{kji} são dados por (3.42) e (3.43), respectivamente, com θ_j substituída por θ_{kj}. Para aplicarmos o método "Scoring"de Fisher, devemos tomar a esperança da expressão acima, resultando em

$$\Delta(\theta_{kj})=-\sum_{i\in\boldsymbol{\mathcal{I}}_k}P_{kji}^*Q_{kji}^*W_{kji}h_{kji}^2-\frac{1}{\sigma_k^2}. \quad (5.22)$$

5.6.3 Estimação por EAP

A estimação de θ_{kj} pela média da posteriori (EAP) consiste em obter a esperança da posteriori, que pode ser escrita como

$$g(\theta|\boldsymbol{u}_{kj.},\boldsymbol{\zeta},\boldsymbol{\eta}_k)=\frac{P(\boldsymbol{u}_{kj.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta}_k)}{P(\boldsymbol{u}_{kj.}|\boldsymbol{\zeta},\boldsymbol{\eta}_k)}. \quad (5.23)$$

Segue que a esperança da posteriori é

$$\widehat{\theta}_{kj} \equiv E[\theta|\boldsymbol{u}_{kj.}, \boldsymbol{\zeta}, \boldsymbol{\eta}_k] = \frac{\int_{I\!R} \theta g(\theta|\boldsymbol{\eta}_k) P(\boldsymbol{u}_{kj.}|\theta, \boldsymbol{\zeta}) d\theta}{\int_{I\!R} g(\theta|\boldsymbol{\eta}_k) P(\boldsymbol{u}_{kj.}|\theta, \boldsymbol{\zeta}) d\theta}. \tag{5.24}$$

Esta forma de estimação tem a vantagem de ser calculada diretamente, não necessitando da aplicação de métodos iterativos. Além disso, as quantidadas necessárias para o seu cálculo são um produto final da etapa de estimação. Por conta disso alguns autores (por exemplo, Mislevy & Stocking (1989)) recomendam esta escolha para a estimação das habilidades.

No próximo capítulo apresentaremos a construção e interpretação da escala de habilidade e uma aplicação prática.

Capítulo **6**

A Escala de Habilidade e uma Aplicação Prática

6.1 Introdução

Neste capítulo vamos descrever os procedimentos para a construção de escalas de habilidade e em seguida iremos ilustrar como é feita sua interpretação através de uma aplicação prática da TRI na área de avaliação da aprendizagem.

6.2 Construção e interpretação de escalas de habilidade

Uma vez que todos os parâmetros dos itens e que todas as habilidades dos respondentes — tanto individuais como populacionais — de todos os grupos avaliados estão numa mesma métrica, ou seja, quando todos os parâmetros envolvidos são comparáveis, pode-se então construir escalas de conhecimento interpretáveis.

Devido à natureza arbitrária das estimativas dos parâmetros dos itens e das habilidades, já comentada anteriormente, sabemos que podemos comparar entre si as habilidades obtidas para os diferentes respondentes, mas que no entanto, elas não possuem "*de per si*" qualquer significado prático em termos pedagógicos. Assim, a menos que efetue-se uma ligação desses valores com os conteúdos envolvidos na avaliação, pode-se dizer apenas que um indivíduo com habilidade 1,80 na escala (0,1) deve possuir um conhecimento muito maior do conteúdo avaliado do que um indivíduo com habilidade -0,50, e também que o primeiro indivíduo tem uma habilidade 1,80 desvios-padrão acima da média da população avaliada enquanto que o segundo tem habilidade 0,50 desvios-

padrão abaixo da média dessa mesma população. Por outro lado, não podemos afirmar nada a respeito do que o indivíduo com habilidade 1,80 sabe a mais do que aquele com habilidade -0,50.

Estes fatos motivaram então a criação de escalas de conhecimento — também chamadas de escalas de habilidade —, que tornam possível a interpretação pedagógica dos valores das habilidades. Essas escalas são definidas por *níveis âncora*, que por sua vez são caracterizados por conjuntos de itens denominados *itens âncora*. Níveis âncora são pontos selecionados pelo analista na escala da habilidade para serem interpretados pedagogicamente. Já os itens âncora são itens selecionados, segundo a definição dada abaixo, para cada um dos níveis âncora.

Definição de item âncora: Considere dois níveis âncora consecutivos Y e Z com $Y < Z$. Dizemos que um determinado item é *âncora para o nível* Z se e somente se as 3 condições abaixo forem satisfeitas simultaneamente:

1. $P(U = 1|\theta = Z) \geq 0,65$ e
2. $P(U = 1|\theta = Y) < 0,50$ e
3. $P(U = 1|\theta = Z) - P(U = 1|\theta = Y) \geq 0,30$

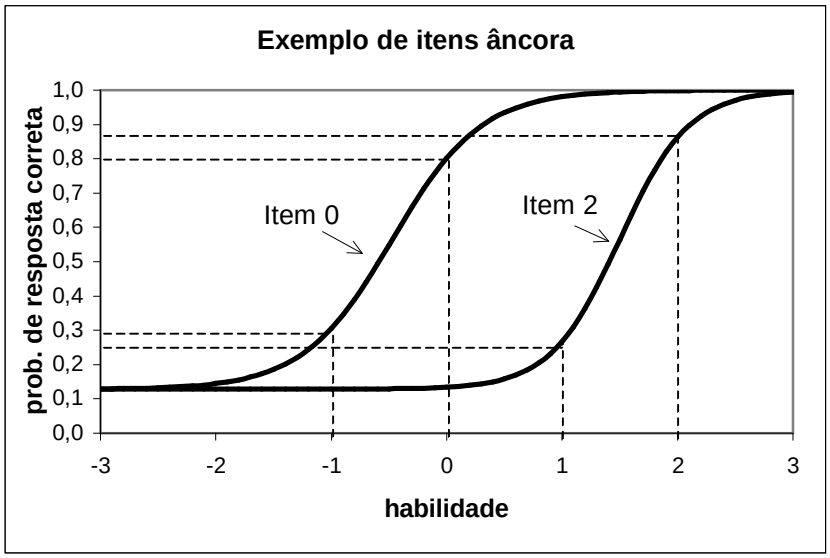
Em outras palavras, para um item ser âncora em um determinado nível âncora da escala, ele precisa ser respondido corretamente por uma grande proporção de indivíduos (pelo menos 65%) com este nível de habilidade e por uma proporção menor de indivíduos (no máximo 50%) com o nível de habilidade imediatamente anterior. Além disso, a diferença entre a proporção de indivíduos com esses níveis de habilidade que acertam a esse item deve ser de pelo menos 30%. Assim, para um item ser âncora ele deve ser um item "típico"daquele nível, ou seja, bastante acertado por indivíduos com aquele nível de habilidade e pouco acertado por indivíduos com um nível de habilidade imediatamente inferior.

Como último comentário, podemos dizer que é bastante comum fazer uma transformação linear em todos os parâmetros envolvidos antes da construção das escalas. Tal procedimento tem como único objetivo facilitar a construção e utilização da escala, uma vez que procura transformar valores negativos ou decimais em números positivos e inteiros.

Na Figura 6.1 são apresentados, em uma escala de habilidade com níveis âncora $-3, -2, -1,\ 0,\ 1,\ 2$ e 3, exemplos de 2 itens âncora (item 0 e item 2) para os níveis âncora 0 e 2, respectivamente. Os parâmetros dos itens são:

$$a_0 = 1,52 \quad , \quad b_0 = -0,47 \quad \text{e} \quad c_0 = 0,13$$
$$a_2 = 1,97 \quad , \quad b_2 = 1,50 \quad \text{e} \quad c_2 = 0,13.$$

Figura 6.1 *Exemplo de 2 itens âncora*



A partir das expressões abaixo, pode-se verificar que os dois itens satisfazem a definição de item âncora:

$(i)\ P(U_0 = 1|\theta = 0) = 0,80 \geq 0,65$
$(ii)\ P(U_0 = 1|\theta = -1) = 0,31 < 0,50$
$(iii)\ P(U_0 = 1|\theta = 0) - P(U_0 = 1|\theta = -1) = 0,80 - 0,31 = 0,49 \geq 0,30$

e

$$(i)\ P(U_2 = 1|\theta = 2) = 0,86 \geq 0,65$$
$$(ii)\ P(U_2 = 1|\theta = 1) = 0,27 < 0,50$$
$$(iii)\ P(U_2 = 1|\theta = 2) - P(U_2 = 1|\theta = 1) = 0,86 - 0,27 = 0,59 \geq 0,30.$$

A priori, não se pode ter certeza de quantos itens âncoras serão selecionados para cada nível âncora e nem se existirão no teste aplicado itens âncoras para todos os níveis âncora determinados. Por isto, é fundamental que os níveis âncoras sejam escolhidos não muito próximos uns dos outros e também que o número de itens aplicados seja bastante grande de modo a possibilitar a construção e interpretação da escala de habilidade. No SAEB por exemplo, foram aplicados 130 itens para cada uma das disciplinas avaliadas na 4.ª série do Ensino Fundamental e 169 itens de cada uma das disciplinas da 8.ª série do Ensino Fundamental e também da 3.ª série do Ensino Médio. Como já foi comentado anteriormente, essa quantidade de itens foi aplicada visando cobrir amplamente a grade curricular de cada uma das séries nas disciplinas avaliadas e também propiciou a identificação e caracterização de diversos níveis âncora para a construção das escalas de habilidades. Maiores detalhes sobre construção e interpretação de escalas de habilidade poderão ser encontrados em Beaton & Allen (1992).

6.3 Uma aplicação prática

A Secretaria de Estado da Educação de São Paulo – SEE/SP implantou, em 1996, o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP, visando alcançar dois objetivos. O primeiro seria ampliar o conhecimento do perfil de realização dos alunos, fornecendo aos professores informações sobre o desempenho dos alunos de modo a subsidiar o trabalho a ser desenvolvido em sala de aula. Assim, os docentes poderiam identificar, no começo do ano escolar, os pontos fortes e fracos do desempenho dos alunos e, a partir desse diagnóstico, adotar estratégias pedagógicas apropriadas.

O segundo seria fornecer informações essenciais para a melhoria da gestão do sistema educacional, na medida em que identifica os pontos críticos do ensino e

possibilita à SEE, por meio de seus órgãos centrais e das Delegacias de Ensino, apoiar as escolas e os educadores com recursos, serviços e orientações.

6.3.1 As características da aplicação

As provas do SARESP são elaboradas a partir de matrizes curriculares, ou seja, tabelas de especificação de conteúdos e objetivos, que indicam os temas e metas do currículo a serem desenvolvidos em cada série e disciplina. Esses parâmetros fundamentam-se nas Propostas Curriculares elaboradas pela Coordenadoria de Estudos e Normas Pedagógicas – CENP e, desde 1997, os itens que compõem as provas vêm sendo construídos pelos professores da Rede Estadual de Ensino.

Até o momento o SARESP foi realizado em 3 anos consecutivos, e a aplicação das provas foi feita segundo a Tabela 6.1.

Tabela 6.1 *Esquema da aplicação das Provas do SARESP*

Ano de Aplicação	Séries e períodos avaliados	Provas compostas pelas Disciplinas
1996	3.ª série diurna do Ensino Fundamental	1-Língua Portuguesa e 2-Matemática
1996	7. ª série diurna e noturna do Ensino Fundamental	1-Língua Portuguesa, 2-Matemática, 3-Ciências e 4-História e Geografia
1997	4.ª série diurna do Ensino Fundamental	1-Língua Portuguesa e 2-Matemática
1997	8.ª série diurna e noturna do Ensino Fundamental	1-Língua Portuguesa, 2-Matemática, 3-Ciências e 4-História e Geografia
1998	5.ª série diurna e noturna do Ensino Fundamental	1-Língua Portuguesa e 2-Matemática
1998	1.ª série diurna e noturna do Ensino Médio	1-Língua Portuguesa, 2-Matemática, 3-Ciências e 4-História e Geografia

Como as avaliações são sempre realizadas no início do ano letivo, as provas de cada uma das séries-alvo são baseadas em conteúdos abordados no ano anterior. Exemplificando, em 1996, as provas dos alunos da 3.^{a} e 7.^{a} séries foram elaboradas com base nos conteúdos relativos ao Ciclo Básico e à 6.^{a} série, respectivamente.

Em todos os anos foram avaliados todos os alunos que frequentavam as séries envolvidas: trata-se, portanto, de uma avaliação de caráter censitário. Cada aluno, entretanto, é avaliado em apenas uma disciplina, ou seja, na 3.^{a} e 4.^{a} séries metade dos alunos responde à prova de Língua Portuguesa e a outra metade, à de Matemática. Essa divisão é feita de maneira aleatória. Nas demais séries, os alunos são divididos, também aleatoriamente, e 25% deles fazem cada uma das 4 provas: Língua Portuguesa, Matemática, Ciências ou História e Geografia. Essa última prova é a única onde aparecem duas disciplinas. No entanto, em termos de análise, as duas disciplinas são obviamente consideradas separadamente.

6.3.2 O tipo de resultados alcançados

A cada ano de aplicação, todos os itens que compõem as provas de cada uma das disciplinas consideradas são cuidadosamente avaliados e interpretados, dentro de cada série envolvida. Para tanto, são considerados tanto seus parâmetros obtidos através da TRI como também algumas estatísticas fornecidas pela Teoria Clássica. A partir dessas informações, um conjunto de especialistas em cada uma das disciplinas faz um diagnóstico completo de cada item (assunto abordado, grau de dificuldade, grau de discriminação, erros mais frequentes, etc.), e também da prova como um todo. Com base nessas informações, pode-se ter uma visão geral do desempenho dos alunos e verificar quais as principais deficiências da série naquele ano.

No entanto, essa avaliação isolada feita ano a ano em cada série, não nos permite comparar o desempenho dos alunos de um ano para o outro, ou seja, verificar se houve realmente um ganho no conhecimento de uma série para a seguinte. Para responder a esta questão, seria necessário que os itens de duas séries consecutivas fossem comparáveis, ou seja, estivessem na mesma métrica. E isto poderia ser conseguido através de uma Equalização.

No entanto, as provas de um ano para outro não apresentavam itens comuns. Como fazer então uma equalização entre duas populações que foram submetidas a provas totalmente diferentes? A solução encontrada foi a criação de uma prova adicional, que serviria de "ligação", uma vez que seria composta de itens que haviam sido submetidos a essas duas populações.

Exemplificando, as provas aplicadas em 1997, na 4.^{a} e 8.^{a} séries, não tinham itens comuns com as provas aplicadas no ano anterior, na 3.^{a} e 7.^{a} séries,

respectivamente. Assim, foram montadas duas provas de ligação: a primeira, composta de itens que haviam sido submetidos à 3.ª série e à 4.ª série e a segunda composta de itens que haviam sido submetidos à 7.ª e à 8.ª séries. Essas duas provas adicionais foram aplicadas no final do ano de 1997, a uma amostra de alunos da 3.ª e da 7.ª séries, respectivamente. Cabe ressaltar, que estes dois grupos adicionais foram introduzidos no estudo com o único objetivo de possibilitar a equalização, não havendo nenhum interesse em estudar o desempenho destas populações.

A partir destas provas de ligação foi possível a criação de uma escala *única* para as séries consecutivas, permitindo assim a comparação dos resultados e a criação de escalas de conhecimento interpretáveis. No SARESP essas escalas foram construídas para as disciplinas Língua Portuguesa e Matemática, por serem as únicas disciplinas avaliadas em todas as séries, todos os anos.

Vamos descrever mais detalhadamente esse processo usando como exemplo as provas de Língua Portuguesa da 3.ª e 4.ª séries.

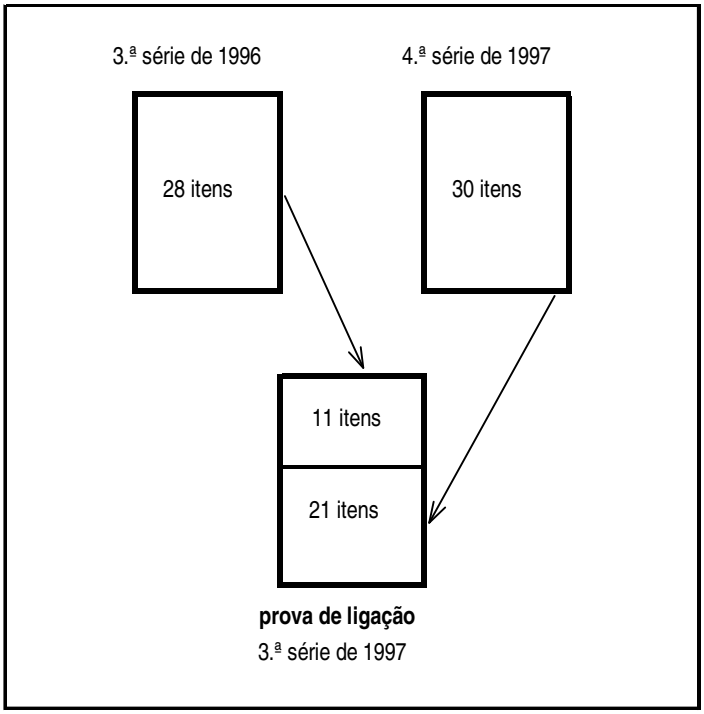
6.3.3 Um exemplo: a Língua Portuguesa na 3.ª e 4.ª séries

Em 1996 foi aplicada uma prova de 28 itens de Língua Portuguesa aos alunos da 4.ª série. Em 1997, os alunos da 4.ª série foram avaliados nessa disciplina através de uma prova composta de 30 itens, totalmente distinta da prova aplicada no ano anterior.

Num primeiro momento, cada uma destas provas teve seus itens calibrados e interpretados dentro de suas respectivas séries. Mas, para que a equalização entre as duas séries pudesse ser possível, foi criada uma prova de ligação, composta de 32 itens, sendo 11 provenientes da prova da 3.ª série e 21 da prova da 4.ª série, como mostra a Figura 6.2. Esta prova foi então submetida a uma amostra de alunos que cursavam a 3.ª série, no final de 1997. Esta nova população foi introduzida no estudo apenas para possibilitar a equalização.

Cabe ressaltar que a prova de ligação foi composta de mais itens da prova da 4.ª série do que da prova da 3.ª, pois houve a preocupação de montar-se uma prova com diferentes graus de dificuldade e com um bom nível de discriminação. Uma vez que as provas de 96 e 97 já haviam sido analisadas separadamente através da TRI, foram selecionados os itens com tais características e a prova da 4.ª série de 97 apresentou um número maior deles. Também é importante notar que a população escolhida para fazer a prova de

Figura 6.2 *Esquema da composição da prova de ligação*



ligação foi a 3.^{a} série de 1997, pois como já foi dito, os itens das provas da 3.^{a} série de 96 e da 4.^{a} série de 97 foram elaboradas com base nos conteúdos dos anos anteriores, ou seja, eram referentes aos conteúdos do Ciclo Básico e da 3.^{a} série, respectivamente. Como a prova de ligação foi aplicada no final do ano letivo de 1997, a série mais indicada para ser submetida a tal prova era, portanto, a 3.^{a} série.

Todos os 58 itens, respondidos pelos alunos das 3 populações envolvidas foram então calibrados simultaneamente, através do modelo de 3 populações discutido no Capítulo 5. Foram utilizados procedimentos bayesianos para a estimação dos parâmetros dos itens e das habilidades. Assim, foram consideradas distribuições a priori para cada um dos parâmetros dos itens e também distribuições normais padrão a priori, para cada uma das populações envolvi-

das. O grupo submetido à prova de ligação (3.ª série de 97) foi considerado a população de referência. Portanto, as outras séries foram posicionadas em relação à ela. No final do processo de estimação, foram fornecidas as estimativas das distribuições a posteriori, para cada uma das populações.

Cabe ressaltar novamente que não havia interesse em estudar o desempenho dos alunos submetidos à prova de ligação, ou seja, ao grupo da 3.ª série de 97. O número de alunos que fizeram essa prova foi apenas o suficiente para atender às exigências da TRI, no que se refere ao número mínimo de sujeitos necessários para obter-se boas estimativas dos parâmetros dos itens. As Figuras 6.2 e 6.3 ilustram a forma dessas distribuições, obtidas para as duas populações de interesse. Para a construção desses gráficos foi utilizada uma amostra de 2059 alunos da 3.ª série de 1996 e 1989 alunos da 4.ª série de 1997.

Figura 6.3 *Representação gráfica da distribuição a posteriori das habilidades em Língua Portuguesa dos alunos da* 3.ª *série*

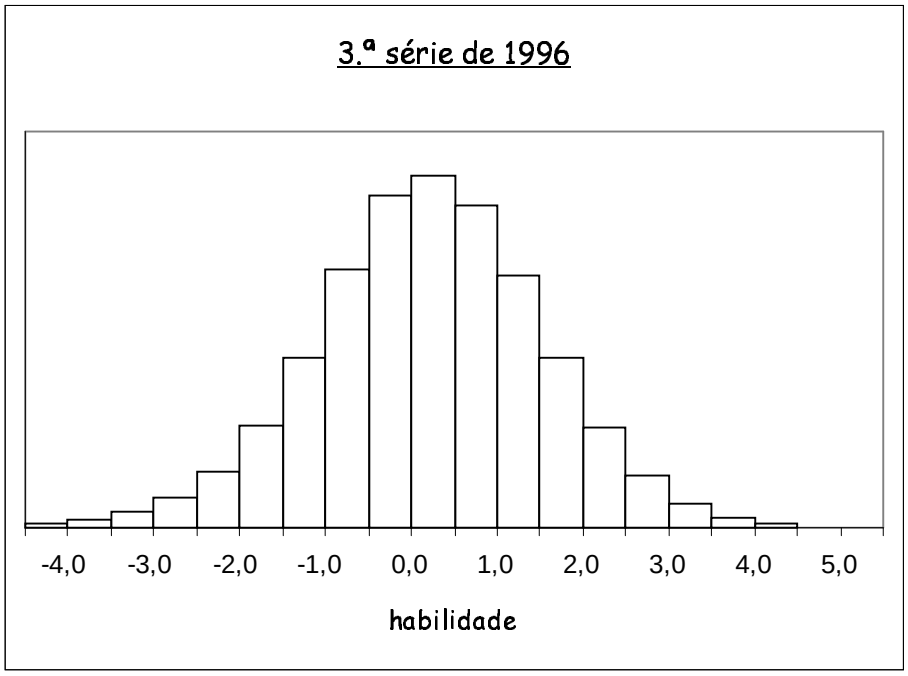
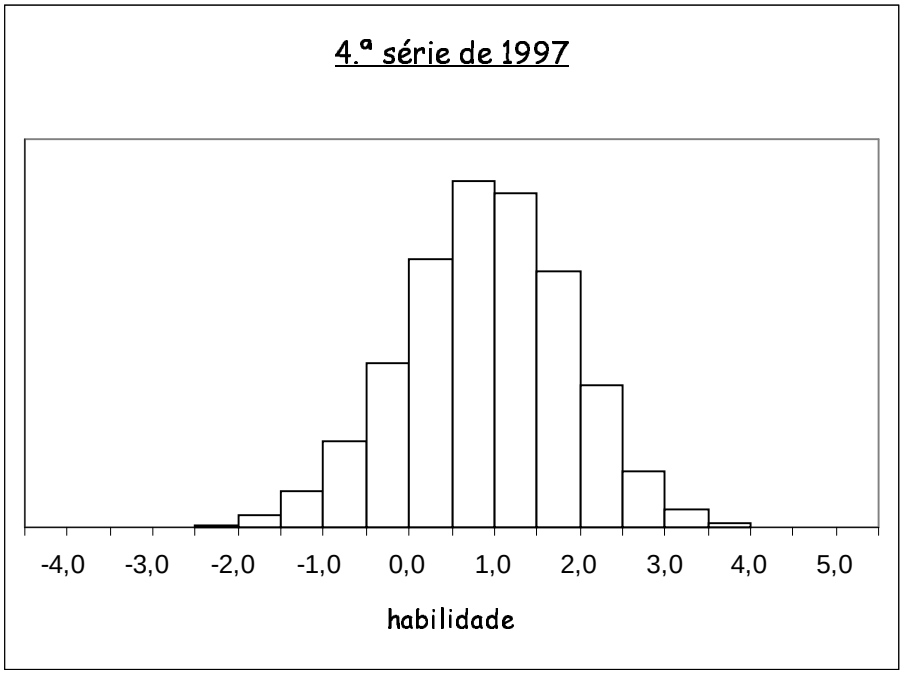


Figura 6.4 *Representação gráfica da distribuição a posteriori das habilidades em Língua Portuguesa dos alunos da 4.ª série*



6.3.4 Interpretação dos resultados

Podemos observar que o gráfico da 4.ª série encontra-se deslocado para a direita, com relação ao gráfico da 3.ª série. Houve um aumento na média da 4.ª série em relação à 3.ª (representada pela linha tracejada). Também podemos observar que os alunos da 4.ª série parecem ser mais homogêneos do que os alunos da série anterior, com relação à habilidade em Língua Portuguesa.

Foi feita uma transformação linear nas estimativas dos parâmetros dos itens e das habilidades dos alunos, visando um melhor entendimento dos resultados. Após essa transformação, a média e o desvio-padrão das habilidades dos alunos da 3.ª série de 1996 em Língua Portuguesa foram fixados em 50 e 16, respectivamente. Para a 4.ª série os valores obtidos foram 62 e 13.

Com todos os 58 itens na mesma métrica, o próximo passo foi a identificação de níveis âncora – conforme descrito na Seção 6.1 – que pudessem caracterizar a escala de conhecimento em Língua Portuguesa da 3.ª e 4.ª séries.

Assim, foi possível a caracterização de 5 níveis âncora (nos pontos 5, 30, 45, 60 e 75) na escala de habilidades de Língua Portuguesa da 3.ª e 4.ª séries. Cada um desses níveis âncora é formado por um conjunto de itens, que caracterizam esse ponto na escala de habilidades, de acordo com a natureza e o grau de conhecimentos que eles exigem.

Após a identificação dos níveis âncora, um grupo de especialistas analisa e interpreta o conjunto de itens que o compõem, a fim de caracterizar cada ponto da escala. A seguir, exemplificamos como ficou a caracterização de um determinado nível âncora da escala de habilidades em Língua Portuguesa da 3.ª e 4.ª séries do SARESP:

`Nível 60 - Língua Portuguesa`

Neste nível, os alunos são capazes de identificar o narrador e revelam ter noções relativas ao papel geral que este assume na história. Com relação ao uso e interpretação da Língua Portuguesa, reconhecem a função do sinal de interrogação no texto. Nos textos narrativos-descritivos, identificam os diferentes elementos que estruturam o texto, discernindo ou reconstituindo a seqüência lógica dos fatos narrados. Em texto de correspondência (bilhete), conseguem interpretar o sentido da mensagem, percebendo implicações lógicas entre as informações contidas no texto.

Demonstram, ainda, certa familiaridade com a leitura de histórias em quadrinhos, fazendo a leitura de imagens e inferindo o significado atribuído a uma expressão onomatopaica como, por exemplo, "PLOFT", *identificado como o barulho de um livro ao ser fechado.*

Além da interpretação de cada ponto que caracteriza a escala de habilidades, também foi calculada a porcentagem de alunos em cada série que dominavam os assuntos descritos em cada nível, visando avaliar os ganhos, em termos de conhecimentos, de um ano para outro. Por exemplo, para o nível 60, descrito anteriormente, chegamos aos seguintes resultados:

Em 1996, a porcentagem de estudantes que respondiam questões desse nível

era de 26,6%. Em 1997, essa porcentagem passa a ser de 55,8%. Ou seja, houve um ganho de 29,2% (pontos percentuais) da 3.^{a} *para a* 4.^{a} *série.*

Por fim, foi estimada a habilidade média (e o respectivo erro-padrão) em Língua Portuguesa, para cada escola. Assim, cada uma delas recebeu um boletim, indicando o desempenho médio da escola, da delegacia da qual ela faz parte e também o resultado médio geral (ou seja, da população toda, que no caso, são todas as escolas públicas estaduais de São Paulo). Com base nessas informações, cada instituição de ensino pode verificar qual sua situação em relação às demais, além de avaliar os ganhos de seus alunos de um ano para outro, e de ter indicações sobre quais os assuntos em que seus alunos ainda estão deficientes.

Obviamente, todos os resultados obtidos são também enviados para as Delegacias de Ensino e para a Secretaria de Estado da Educação de São Paulo. Assim, a partir das informações fornecidas pelo SARESP, as ações podem ser tomadas tanto a nível de cada instituição de ensino, quanto em proporções estaduais.

Dando prosseguimento ao estudo, em 1998 uma das séries avaliadas pelo SARESP foi a 5.^{a} série do Ensino Fundamental, nos períodos diurno e noturno. Para cada uma das disciplinas avaliadas dois tipos de provas, com alguns itens comuns, foram aplicados em cada uma das populações — diurna e noturna. Novamente, as provas aplicadas não tinham itens comuns com as provas dos anos anteriores.

Mais uma vez, foi montada uma prova de ligação, composta de itens utilizados nas provas de 3 das 4 populações de interesse: 4.^{a} série de 1997, 5.^{a} série diurna de 1998 e 5.^{a} série noturna de 1998. Essa prova foi aplicada então a uma amostra de alunos que cursavam a 4.^{a} série em 1998. Essa população adicional também foi introduzida no estudo apenas com o objetivo de possibilitar a equalização.

Cabe ressaltar que a meta agora era colocar os alunos da 3.^{a} série de 96, 4.^{a} série de 1997 e 5.^{a} séries diurna e noturna de 98, todos na mesma escala. Nessa nova equalização, os itens da 3.^{a} série não precisaram mais entrar na prova de ligação, pois a 3.^{a} e a 4.^{a} séries já haviam sido colocadas na mesma métrica. Na verdade, agora é como se fossemos apenas "colar"a 5.^{a} série nas séries anteriores. Assim, essa segunda equalização foi realizada de uma maneira bastante distinta da primeira. Os itens calibrados no ano anterior foram

mantidos fixos durante o processo de estimação e apenas os itens aplicados à 5.^{a} série foram calibrados, resultando ao final do processo, num conjunto de itens de 3.^{a} à 5.^{a} séries, todos na mesma escala. Dessa maneira, a escala de habilidades da 3.^{a} e da 4.^{a} séries pode ser ampliada com a entrada da 5.^{a} série e interpretada para todo esse conjunto de alunos.

Concluindo, esse estudo, além de avaliar o desempenho da rede estadual de São Paulo ano a ano, também vem fornecendo indicadores quantitativos de como as intervenções no ensino público têm afetado o conhecimento dos alunos de uma série para outra, e esse tipo de questão só pode ser respondida através das ferramentas fornecidas pela TRI.

No próximo capítulo, discutiremos alguns dos recursos computacionais disponíveis para a análise de dados via TRI. Em particular, descreveremos o desempenho de dois programas computacionais frente aos diferentes tipos de equalização abordados no Capítulo 4.

Capítulo **7**

Recursos computacionais

7.1 Introdução

Sem dúvida alguma, o crescimento e a divulgação da TRI sempre estiveram intimamente ligados ao desenvolvimento paralelo de recursos computacionais que viabilizassem sua utilização. Isto porque as ferramentas matemáticas necessárias para sua aplicação são muito mais complexas do que as técnicas empregadas na Teoria Clássica de Medidas.

Desde suas primeiras aplicações, pesquisadores têm desenvolvido seus próprios programas computacionais, mas é certo que sua utilização em larga escala depende diretamente da disponibilidade de programas computacionais comerciais no mercado. Na Europa e nos Estados Unidos, desde a década de 70 foram lançados vários programas específicos para análise via TRI. Aqui no Brasil, onde a utilização da TRI é bem mais recente, há uma variedade bem menor de programas computacionais comerciais sendo usados.

Neste capítulo, vamos comentar os programas computacionais comerciais mais usados atualmente no Brasil e que se propõem a resolver, na prática, muitos dos problemas abordados pela TRI e que foram descritos nos capítulos anteriores.

7.2 Recursos computacionais

Iniciaremos pelo programa TESTFACT (ver Wilson et al.(1991)) que produz várias estatísticas descritivas para os itens de um teste, inclusive algumas das utilizadas pela teoria clássica, mas que também tem recursos importantes para a TRI, usados na verificação da dimensionalidade dos testes: técnicas de análise fatorial específicas para serem aplicadas em itens. Dois tipos especiais de análise fatorial, que foram elaboradas para variáveis dicotômicas (como é

o caso dos itens, quando são considerados como certo ou errado), estão implementados neste programa. Uma delas é a análise fatorial feita a partir da matriz de correlação tetracórica, que é um tipo especial de correlação, utilizada quando as variáveis assumem apenas os valores 0 ou 1 (ver Divgi (1979)). A outra técnica implementada é a análise fatorial plena, baseada no método de máxima verossimilhança (ver Mislevy (1986b)).

Para a análise de itens não dicotômicos, podemos citar o programa PARSCALE (ver Muraki & Bock (1997)), que tem implementados os modelos de Resposta Gradual e de Créditos Parciais, descritos no Capítulo 2. Em sua versão mais recente, é possível fazer análises para mais de um grupo de respondentes.

Dos programas disponíveis no mercado, os que são atualmente mais utilizados nas análises envolvendo a TRI - aqui no Brasil - são o BILOG (ver Mislevy & Bock (1990) e o BILOG-MG (ver Zimowski et al. (1996)). Estes dois programas são específicos para análises via TRI de itens dicotômicos ou dicotomizados e ambos têm implementados os modelos unidimensionais logísticos de 1, 2 e 3 parâmetros. A diferença básica entre eles é que o BILOG-MG permite a análise de mais de um grupo de respondentes, enquanto que o BILOG permite apenas analisar respondentes considerados como provenientes de uma única população.

Vamos comentar a seguir quais dos métodos de estimação descritos nos Capítulos 3 e 5 estão implementados nestes dois programas e também dar uma ênfase especial ao desempenho deles perante as diversas situações que envolvem equalizações, descritas no Capítulo 4.

7.2.1 Os programas BILOG for Windows v. 3.09 e BILOG-MG v. 1.0

Esses dois programas executam a análise em três etapas, chamadas de fases 1, 2 e 3, que se caracterizam pelo tipo de tarefas realizadas em cada uma delas. Na fase 1, que é a fase de entrada e leitura de dados, o usuário deve fornecer ao programa basicamente dois tipos de informação: a identificação de cada indivíduo com suas respectivas respostas ao teste e o gabarito (que é uma sequência contendo as alternativas corretas dos itens que compõem o teste). Também é possível fornecer as respostas já corrigidas, ou seja, já codificadas como 0 ou 1. Nesse caso não há a necessidade do gabarito, pois o

programa irá interpretar 1 como acerto e 0 como erro. No caso de esquemas amostrais complexos, pode-se fornecer ao programa pesos diferentes para cada um dos indivíduos. Essas informações devem estar em arquivos do tipo ASCII. Os arquivos de saída, fornecidos ao usuário, também estarão neste formato. Nessa fase é feita a"correção"da prova de cada respondente (no caso de ter sido fornecido o arquivo com as respostas originais) e são calculadas algumas estatísticas descritivas, tais como: número de indivíduos submetidos a cada item, número e porcentagem de acerto em cada item e algumas correlações de interesse, como as correlações bisserial e ponto bisserial (ver Lord & Novick (1968), por exemplo), usadas na Teoria Clássica de Medida. A importância dessa etapa do processamento, além da verificação de que a leitura dos dados foi feita corretamente, é que estas estatísticas são utilizadas posteriormente como valores iniciais para os processos de estimação realizados nas fases seguintes. Além disso, estatísticas como a correlação bisserial, fornecem um diagnóstico preliminar dos itens, servindo por exemplo, na identificação de itens com problemas no gabarito.

A fase 2 é a fase da calibração dos itens. Nesta fase, são estimados os parâmetros dos itens, com seus respectivos erros-padrão. Os métodos de estimação disponíveis serão comentados na próxima seção. O BILOG fornece ainda gráficos contendo algumas informações de interesse, tais como as curvas características e as curvas de informação de cada item e do teste. No BILOG-MG esses gráficos também podem ser obtidos, mas com uma resolução bastante baixa. Isto se deve ao fato de que o programa BILOG já está disponível para o sistema Windows, enquanto que o BILOG-MG ainda só tem versões para o sistema operacional DOS. Junto com a curva característica de cada item é fornecido também um teste de ajuste do modelo utilizado.

A fase 3 é a fase da estimação das habilidades dos respondentes. Aqui são estimadas as habilidades de cada um dos indivíduos, a partir dos resultados obtidos na fase anterior. Essas habilidades inicialmente são estimadas na escala dos parâmetros dos itens. No entanto, pode-se especificar alguns tipos de mudanças na escala, que serão feitas tanto nas habilidades como nos parâmetros estimados na fase anterior. Maiores detalhes quanto aos métodos de estimação realizados nesta fase que estão disponíveis nesses programas serão fornecidos na próxima seção.

7.2.2 Métodos para a calibração dos itens

Como foi dito na seção anterior, esses dois programas realizam inicialmente a calibração (estimação dos parâmetros) dos itens e depois a estimação das habilidades dos respondentes. Dois métodos de estimação para os parâmetros dos itens estão implementados, tanto no BILOG, como no BILOG-MG: máxima verossimilhança marginal e um método bayesiano de estimação por maximização da distribuição marginal a posteriori.

Assim, como foi descrito nos Capítulos 3 e 5, para que os parâmetros dos itens possam ser estimados através de qualquer um desses dois métodos, é necessária a utilização de distribuições de probabilidade para as habilidades dos respondentes. Esses programas assumem que os respondentes representam uma amostra aleatória de uma população de habilidades que pode ser assumida como tendo ou uma distribuição normal padrão, ou uma distribuição discreta arbitrariamente especificada pelo usuário, ou ainda uma distribuição empírica, a ser estimada conjuntamente com os parâmetros dos itens. Esta distribuição empírica é representada na forma de uma distribuição discreta, através de pontos de quadratura.

No caso de mais de um grupo de respondentes, quando usamos o BILOG-MG, ao final do processo de calibração dos itens são fornecidas também estimativas da média e desvio-padrão da distribuição de habilidades a posteriori para cada população.

Também, como já foi citado nos capítulos sobre estimação, na estimação por maximização da distribuição marginal a posteriori, distribuições a priori são definidas para os parâmetros dos itens. No caso desses dois programas, o usuário pode especificar prioris normais para o parâmetro de dificuldade, prioris log-normais para os parâmetros de discriminação e prioris beta para o parâmetro de acerto casual.

O BILOG e o BILOG-MG utilizam duas formas de resolver as equações de verossimilhança marginal: o algoritmo EM e o método "Scoring"de Fisher.

7.2.3 Métodos implementados para a estimação das habilidades

Uma vez terminada a calibração dos parâmetros, será feita a estimação das habilidades dos respondentes. O BILOG e o BILOG-MG têm implementados os métodos de estimação por máxima verossimilhança, por esperança a

posteriori (EAP) e por máximo a posteriori (MAP). No método da máxima verossimilhança, as estimativas das habilidades dos respondentes são calculadas pelo método de Newton-Raphson, utilizando-se uma transformação linear do logito do percentual de acertos dos indivíduos como valores iniciais. Os problemas já descritos com as estimativas dos respondentes que tiveram erro total ou acerto total são contornados através de um artifício: os alunos que erraram todos os itens ganham um meio certo no item mais fácil. Alunos que acertaram todos os itens, perdem um meio certo no item mais difícil. Apesar dessas alternativas implementadas pelos dois programas, este método nem sempre fornece boas estimativas nestes casos. No método EAP, as estimativas para as habilidades são calculadas utilizando-se pontos de quadratura para aproximar a distribuição a priori das habilidades de cada respondente. O número de pontos de quadratura é definido pelo usuário, que pode também escolher entre uma priori que seja normal (e cujos parâmetros podem ser especificados pelo usuário), ou uma distribuição discreta arbitrária (fornecida pelo usuário), ou ainda uma distribuição discreta empírica, através do uso dos pontos de quadratura e de seus respectivos pesos gerados na fase 2. As estimativas EAP para as habilidades dos respondentes estão sempre definidas, qualquer que seja o padrão de respostas. Além disso, quando utilizamos a estimação por EAP, é fornecida uma estimativa da distribuição de habilidades da população de respondentes, na forma de uma distribuição discreta, dada pelos pontos de quadratura. Esta distribuição é obtida acumulando-se as densidades a posteriori de todos os sujeitos em cada ponto de quadratura. As somas são então normalizadas para obter-se as probabilidades estimadas em cada ponto. Também são fornecidos a média e o desvio-padrão para essa distribuição estimada. No método MAP, as estimativas das habilidades são calculadas pelo método de Newton-Gauss. Este procedimento sempre converge e fornece estimativas para todos os padrões de resposta possíveis. É assumida uma distribuição a priori normal, cujos parâmetros podem ser especificados pelo usuário, sendo que o padrão definido nesses programas é a normal padrão.

7.3 A equalização nos programas BILOG e BILOG-MG

Quando desejamos que a equalização seja feita durante o processo de calibração dos itens, o uso de programas computacionais especificamente desenvolvidos para esse fim são uma ferramenta bastante importante. O BILOG-MG é um bom exemplo de um programa que pode ser utilizado na maioria dos casos descritos no Capítulo 4. Nesta seção, vamos então descrever quando é possível sua utilização em cada um daqueles casos e em quais deles o BILOG também pode ser usado. Os casos 1 a 6 tratam, respectivamente, das situações descritas nas Seções 4.2.1 a 4.2.6 do Capítulo 4. Já os casos (a) a (c) tratam, respectivamente, das situações descritas nas Seções 4.3.1 a 4.3.3.

7.3.1 O BILOG e o BILOG-MG frente a populações e/ou provas distintas

Caso 1: Aqui temos um único grupo fazendo uma única prova. Por se tratar do caso mais básico, em que não se faz necessário nenhum tipo de equalização, podemos utilizar qualquer um dos programas computacionais disponíveis que tratam da TRI, inclusive o BILOG e o BILOG-MG.

Caso 2: Aqui temos um único grupo fazendo duas provas totalmente diferentes. Por se tratar de um caso de equalização via população, basta que todos os itens de ambas as provas sejam calibrados simultaneamente. Para tanto, devemos fazer apenas uma ligeira alteração nos modelos já propostos, incorporando a informação da prova a que cada aluno foi submetido, uma vez que a cada prova está associado um conjunto de itens distintos. Este também é um caso bastante comum que a maioria dos programas computacionais para análise via TRI é capaz de resolver. No BILOG-MG está situação é tratada sem maiores problemas. Já no BILOG, há uma limitação técnica: as respostas dos alunos devem estar já corrigidas (codificadas com 0 ou 1), para que não haja necessidade de utilizar os gabaritos das provas, uma vez que o programa não consegue ler 2 tipos de gabaritos distintos.

Caso 3: Aqui temos um único grupo fazendo duas provas parcialmente diferentes, isto é, com alguns itens comuns. Este caso é bastante semelhante ao caso anterior, ou seja, a equalização também pode ser feita via população. A única observação que podemos acrescentar é que devemos ter bastante cuidado no tratamento dos itens comuns. É que embora esses itens apareçam nas duas provas, eles não podem ser"contados"duas vezes, ou seja, o número total de itens a ser calibrado é o total de itens da prova A, mais o total de itens da prova B, menos o número de itens comuns entre A e B.

Caso 4: Aqui temos dois grupos fazendo uma mesma prova. Por se tratar de uma situação onde se faz necessária uma equalização via itens comuns, este caso necessita de programas computacionais para análise via TRI que tenham implementados modelos para mais de um grupo. O BILOG, por exemplo, não comporta esse tipo de problema, enquanto que o BILOG-MG foi especialmente desenvolvido para modelar esse tipo de situação. Se só dispuséssemos do BILOG, uma alternativa seria calibrar as provas dos dois grupos separadamente, e depois realizar uma equalização a posteriori, como foi descrito no Capítulo 4. Nesse caso, como todos os itens são comuns, métodos de equalização a posteriori, como o método Média-Desvio, produzem resultados bastante satisfatórios, quando comparados à equalização feita durante o processo de calibração dos itens (ver Andrade (1999), por exemplo).

Caso 5: Aqui temos dois grupos fazendo duas provas totalmente diferentes. Como já foi explicado no Capítulo 4, não há nenhuma maneira de tornar comparáveis os resultados desses dois grupos.

Caso 6: Aqui dois grupos são submetidos a duas provas diferentes, mas que têm alguns itens comuns. Assim como no Caso 4, esta é uma situação típica para ser abordada no BILOG-MG, utilizando-se um modelo para mais de uma população e, portanto, não é possível o uso do BILOG. Como já foi citado no caso 3, devemos apenas ter o cuidado de não considerar duas vezes os itens repetidos. Assim como foi comentado no Caso 4, aqui também pode-se resolver o problema através de uma equalização a posteriori. No entanto, o desempenho desse tipo de equalização torna-se bastante inferior à equalização feita durante o processo de calibração se o número de itens comuns for pequeno.

7.3.2 O BILOG e o BILOG-MG frente ao conjunto de itens a ser calibrado

Caso (a): todos os itens são"novos". Quando desejamos calibrar o conjunto completo de itens, temos o problema de estimação mais comum e, portanto, ele pode ser resolvido utilizando-se qualquer um dos programas computacionais disponíveis que tratam da TRI, inclusive o BILOG e o BILOG-MG.

Caso (b): todos os itens já são calibrados. Se não desejamos calibrar nenhum dos itens, estamos interessados apenas em estimar as habilidades dos respondentes. Este problema pode ser resolvido de maneira relativamente simples através dos programas BILOG e BILOG-MG, sendo necessário apenas fornecermos um arquivo contendo as estimativas dos parâmetros de interesse. No entanto, cabe aqui uma observação: quando se utilizar o BILOG ou o BILOG-MG é sempre recomendável utilizar as estimativas dos parâmetros na escala " original", isto é, como foram fornecidas pelo programa, sem que tenham sofrido nenhum tipo de transformação linear. Isto porque, quando se utiliza os métodos EAP ou MAP para estimar as habilidades dos respondentes, faz-se necessário o uso de uma distribuição a priori para a habilidade de cada um desses respondentes, e o padrão desses programas é utilizar a distribuição normal padrão ou outras distribuições discretas, mas sempre com média e desvio-padrão nas vizinhanças dos valores 0 e 1, respectivamente. Assim, se por exemplo, a métrica da população em que os parâmetros foram estimados tiver sido transformada para (200,40), haverá problemas na estimação das habilidades dos novos respondentes.

Caso (c): alguns itens são"novos"e outros já estão calibrados. Neste caso, desejamos calibrar alguns itens e manter os parâmetros de outros, que já foram calibrados anteriormente. Para que possamos fixar parâmetros de alguns itens e calibrar o restante utilizando o BILOG e o BILOG-MG, deveremos necessariamente utilizar um método de estimação bayesiano, uma vez que o único procedimento disponível nesses programas para fixar apenas parte dos itens, é o uso de distribuições a priori convenientes para os parâmetros desses itens. Para os itens novos, que desejamos calibrar, utilizamos as prioris padrão sugeridas pelo programa. Já para os outros itens, definimos prioris cujas médias

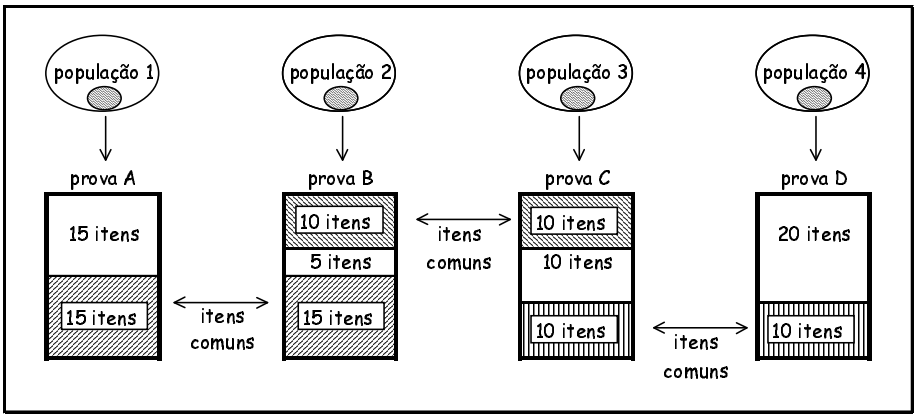
são os próprios valores dos parâmetros que desejamos fixar e cujos desvios-padrão são tão pequenos que a distribuição torna-se praticamente degenerada naquele ponto. O que ocorre na prática é que todos os parâmetros são estimados novamente, mas a convergência daqueles itens conhecidos é artificialmente induzida para os valores que desejamos. Pode-se também "reforçar"ainda mais a convergência utilizando-se outro recurso do programa, que é a definição, por parte do usuário, de valores iniciais convenientes. Mas, o uso deste tipo de procedimento pode acarretar alguns problemas. Por exemplo, se não utilizarmos novamente o mesmo grupo de respondentes da calibração inicial, poderemos ter problemas para obter a convergência nessa segunda calibração. E, na prática, muitas vezes não dispomos do conjunto original de respondentes para juntarmos aos respondentes da nova aplicação. E devemos ressaltar que estamos nos referindo ao caso em que há uma única população sendo submetida a uma única prova. O problema se torna ainda mais complexo, no caso de termos mais de uma população envolvida (comentaremos essa situação a seguir).

7.3.3 O uso do BILOG-MG quando desejamos fixar parte dos itens e calibrar o restante, e há mais de uma população envolvida

Quando há duas (ou mais) populações envolvidas (Casos 4 e 6), e utilizamos o BILOG-MG para estimar parte do conjunto de itens, fixando os demais (Caso (c)), poderemos ter problemas com a métrica. É que, como há mais de uma população envolvida nos processos de estimação, para resolver os problemas de indeterminação de escala, o programa pede ao usuário que defina uma das populações como sendo a referência, que será definida como tendo média 0 e desvio-padrão 1, e então, as demais populações serão posicionadas com relação à ela. Vamos então imaginar a seguinte situação, ilustrada na Figura 7.1: utilizamos amostras das populações 1 e 2 para calibrar um conjunto de itens, provenientes de duas provas (A e B). Estas provas tinham 30 itens cada, sendo 15 itens comuns. A população 1 foi utilizada como referência. Ao final do processo, temos um conjunto de 45 itens (= 30 + 30 - 15) calibrados, além das habilidades dos respondentes das duas populações. Digamos que as estimativas obtidas para os parâmetros populacionais dos dois grupos tenham sido, respectivamente, (0,1) e (2,2). Desse modo, um item i, cuja estimativa do parâmetro b foi 1 está, usando-se como unidade o desvio-padrão da po-

pulação 1, 1 desvio-padrão acima da média da população 1 (e portanto, é relativamente difícil para este grupo) e 1 desvio-padrão abaixo da média da população 2 (e portanto, é relativamente fácil para este grupo). Suponha agora que temos outras duas provas C e D, que serão submetidas, respectivamente, a amostras das populações 3 e 4. Ambas as provas são compostas de 30 itens, sendo que há 10 itens comuns entre elas. Suponha ainda que além disso, há 10 itens na prova C que são comuns com a prova B e, portanto, que já foram calibrados anteriormente.

Figura 7.1 *Esquematização dos itens comuns entre as provas*



Desejamos então fixar os parâmetros desses 10 itens obtidos na calibração anterior e estimar todos os restantes. O motivo para isto seria que, procedendo desta maneira, faríamos uma equalização entre as populações 1, 2, 3 e 4, tornando possível qualquer comparação entre elas. Mas, o que aconteceria se, para tanto, utilizássemos apenas as populações 3 e 4? Para começar, teríamos que definir uma população de referência, digamos a população 3. Logo, essa população será definida como tendo parâmetros (0,1), para que a população 4 seja posicionada com relação a ela. Supondo que aquele item i, cujo valor de b é 1, foi um dos 10 itens que tiveram seus parâmetros fixados, que interpretação deveríamos ter sobre a relação desse item com a população 3? A mesma que já tivemos com relação à população 1: que ele está 1 desvio-padrão acima da média da população 3 e portanto, é relativamente difícil para este grupo. O

fato de termos as populações 1 e 3 necessariamente com a mesma distribuição de probabilidade é um problema, pois sabemos que se tratam de populações diferentes. Suponhamos que essas populações sejam, respectivamente, a 3.ª, a 4.ª, a 5.ª e a 6.ª séries do ensino fundamental. Seria perfeitamente razoável esperarmos que as médias das distribuições de habilidades destas populações mantivessem uma relação crescente de ordem. Assim, se a 3.ª série fosse fixada como tendo parâmetros (0,1) e a 4.a série tivesse então seus parâmetros estimados em (2,2), esperaríamos ter uma média maior do que 2 para a 5.ª série, e não (0,1). Desta maneira, aquele item i, cujo parâmetro de dificuldade foi estimado em 1, deveria estar necessariamente abaixo da média da 5.ª série. Há pelo menos 2 maneiras de solucionarmos este problema. A primeira, que nem sempre é possível, é utilizarmos novamente os respondentes utilizados nas provas A e B no processo da calibração das provas C e D. Fixaríamos todos os itens das provas A e B enquanto calibraríamos os itens novos das provas C e D. Desta maneira, poderíamos definir novamente a população 1 como sendo a referência, e então não haveriam mais problemas no posicionamento das populações 3 e 4. Mas, como já foi dito, nem sempre é possível proceder desta maneira, pois poderíamos não dispor dos respondentes utilizados na primeira calibração. Uma outra maneira de solucionar o problema de maneira adequada, seria fazer uma equalização a posteriori, que já foi comentada na Seção 4.4.

No próximo capítulo serão feitas considerações finais sobre esse trabalho e algumas sugestões para futuras pesquisas e aplicações.

Capítulo **8**

Considerações gerais

Para finalizar, faremos uma breve discussão sobre os problemas encontrados na aplicação dessa teoria, possíveis tópicos de pesquisa e a utilização da TRI em outras áreas do conhecimento.

Nesse livro procuramos introduzir os principais conceitos, modelos, métodos de estimação e aplicações da Teoria da Resposta ao Item, com o objetivo de mostrar o grande potencial da sua aplicação na área de avaliação educacional, em especial quando há a necessidade da comparação do desempenho de duas ou mais populações de indivíduos.

Apesar desta teoria ter mais de 50 anos, somente nos últimos 15 é que ela vem sendo aplicada em larga escala nas principais avaliações educacionais de diferentes países. Atribui-se este fato à complexidade matemática dos métodos envolvidos, praticamente inviáveis sem o auxílio do computador. O que temos observado é que a teoria vem sendo desenvolvida num ritmo que ainda não vem sendo acompanhado pelo desenvolvimento de programas computacionais eficientes, que viabilizem sua utilização em maior escala. Além disso, a aplicação apropriada desta teoria exige necessariamente o envolvimento de especialistas em avaliação e em estatística. Nesse sentido, faz-se imprescindível a elaboração de grupos de trabalho, que possibilitem a integração de profissionais de ambas as áreas. Justamente pelo fato da TRI ter sido ainda tão pouco explorada, vários pontos têm sido levantados na literatura sobre sua adequação. Alguns deles ainda permanecem em aberto.

Podemos citar, por exemplo, a questão da dimensionalidade do espaço de traços latentes envolvidos na avaliação. Todos os modelos que vêm sendo efetivamente utilizados pressupõem que o conhecimento que se deseja medir pode ser representado por uma única habilidade. Alguns autores têm defendido a tese de que os modelos unidimensionais têm fornecido bons resultados, mesmo em situações multidimensionais, desde que uma das dimensões possa ser con-

siderada predominante. Mais recentemente, modelos para mais de uma dimensão têm sido propostos, mas ainda não têm sido aplicados devido a não disponibilidade de recursos computacionais e também à sua maior dificuldade de interpretação. Um estudo interessante seria o da dimensionalidade da prova objetiva do Exame Nacional do Ensino Médio (ENEM), cujos itens são elaborados a partir de situações-problema devidamente contextualizadas na interdisciplinaridade das ciências e das artes em sua articulação com o mundo em que vivemos.

A questão da equalização entre diferentes populações também sempre foi um ponto bastante discutido na literatura. Conforme comentamos neste trabalho, a proposta recente de modelos para vários grupos de Bock & Zimowski (1997), que viabilizam a equalização durante o processo de calibração, deu um novo rumo à solução desta questão, tendo em vista que os modelos anteriores envolvem outros erros de modelagem, além daqueles da própria teoria. Sugerimos a leitura de Goldstein & Wood (1989), Mislevy (1992), Goldstein (1994) e Hedges & Vevea (1997), entre outros, para um melhor entendimento destes problemas e suas soluções.

Outro ponto que poderíamos citar, foi levantado por Mislevy (1991) e diz respeito à qualidade da estimação da distribuição das habilidades dos elementos de uma população. O autor discute a possibilidade de se obter melhores estimativas da variabilidade das habilidades, utilizando-se também outras informações dos respondentes que possam estar associadas com suas habilidades. Exemplos dessas informações seriam o grau de escolaridade dos pais, o hábito de leitura do respondente, a condição sócio-econômica da família, etc. Esta metodologia é baseada no conceito de imputação múltipla de dados faltantes e os valores obtidos para as habilidades são denominados de"valores plausíveis". Mas, ainda existem alguns fatores que dificultam a aplicação desta metodologia, e o principal deles como sempre, é a não existência comercial, até o presente momento, de programas computacionais apropriados. Além disso, há também a dificuldade da obtenção de informações adicionais relevantes ao problema que sejam fidedignas e a inclusão dessas mesmas informações no modelo.

Há ainda outros pontos que têm sido poucos explorados, como por exemplo, modelos multivariados e modelos longitudinais. Os modelos multivariados seriam adequados para as situações onde um mesmo respondente é submetido a mais de um teste e os modelos longitudinais, para as situações onde o desem-

penho de um mesmo respondente é acompanhado ao longo do tempo. Esses últimos modelos deveriam permitir a incorporação de possíveis estruturas de covariância entre as habilidades dos indivíduos avaliados ao longo do tempo. Estes modelos poderiam ser aplicados, por exemplo, nas análises dos dados gerados pelo projeto AVEJU, da Secretaria de Estado da Educação do Estado de São Paulo, que acompanhou um grupo de alunos da escola pública estadual da 1a. série (1992) até a 3a. série (1994) do Ensino Fundamental, e do projeto FUNDESCOLA em implementação pelo INEP/MEC, que deverá acompanhar um grupo de alunos de escolas públicas de 6 estados, desde a 4a. série (1999) até a 8a. série (2003) do Ensino Fundamental.

Para finalizar, gostaríamos de ressaltar dois outros pontos. O primeiro diz respeito a disseminação do uso da TRI em avaliações educacionais brasileiras, que sem dúvida alguma dependerá muito da integração de especialistas das áreas de estatística e educação. A criação de programas de pós-graduação envolvendo departamentos de estatística e de medidas em educação em algumas de nossas universidades, seria de fundamental importância. A primeira aplicação da TRI no Brasil foi na análise do SAEB 95. Desde então, os órgãos governamentais, através do MEC e algumas Secretarias da Educação, vem valorizando e incentivando o uso dessa teoria nas suas avaliações. No entanto, o mercado de trabalho ainda está bastante deficiente de profissionais com tais qualificações. O segundo ponto diz respeito a disseminação do uso da TRI em outras áreas do conhecimento.

Um ponto importante dessa metodologia é que tanto os itens, através de seus parâmetros, quanto o traço latente associado são medidos em uma mesma métrica, permitindo com isso uma operacionalização dessa característica latente que está sendo medida, bem como a adequação e a contribuição de cada um dos itens aplicados nessa operacionalização. Essa propriedade tem levado pesquisadores de diferentes áreas a aplicarem o modelo de Rasch, modelo com um único parâmetro (o parâmetro b de dificuldade), na análise e interpretação de vários instrumentos de avaliação (medida). Três exemplos recentes seriam os trabalhos de DeRoos & Allen-Meares (1998), Tennant et. al. (1996) e Granger et. al. (1998). O primeiro em psiquiatria e os dois últimos em reabilitação médica. O modelo de Rasch é também descrito em Marcoulides (1998) como um dos métodos modernos mais importantes para a pesquisa na área de negócios. Sugerimos aos leitores mais interessados a participação nas

listas de discussão *rasch@acer.edu.au* e *irt@listserv.vt.edu* e uma pesquisa no site *http://www.rasch.org/*.

Apêndice **A**

Desenvolvimento das expressões do Capítulo 3

A.1 Expressões da página 40

Para chegarmos às expressões de $\boldsymbol{H}(\boldsymbol{\zeta}_i)$ e $\boldsymbol{h}(\boldsymbol{\zeta}_i)$ usadas em (3.25), notemos que de (3.22),

$$\frac{\partial \log L(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} = \sum_{j=1}^{n} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right) \left(\frac{\partial^2 P_{ji}}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \right) - \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right) \left(\frac{\partial P_{ji}}{\partial \boldsymbol{\zeta}_i} \right)' \right\}. \tag{A.1}$$

Porém,

$$\begin{aligned}
\frac{\partial P_{ji}^* Q_{ji}^*}{\partial \alpha_i} &= (1 - 2P_{ji}^*) \frac{\partial P_{ji}^*}{\partial \alpha_i}, \quad \alpha_i \in \{a_i, b_i, c_i\}, \\
\frac{\partial^2 P_{ji}}{\partial a_i^2} &= D^2 (1 - c_i)(\theta_j - b_i)^2 P_{ji}^* Q_{ji}^* (1 - 2P_{ji}^*), && \text{(A.2)} \\
\frac{\partial^2 P_{ji}}{\partial a_i \partial b_i} &= -D(1 - c_i) P_{ji}^* Q_{ji}^* \{1 + Da_i(\theta_j - b_i)(1 - 2P_{ji}^*)\}, && \text{(A.3)} \\
\frac{\partial^2 P_{ji}}{\partial a_i \partial c_i} &= -D(\theta_j - b_i) P_{ji}^* Q_{ji}^*, && \text{(A.4)} \\
\frac{\partial^2 P_{ji}}{\partial b_i^2} &= D^2 a_i^2 (1 - c_i) P_{ji}^* Q_{ji}^* (1 - 2P_{ji}^*), && \text{(A.5)} \\
\frac{\partial^2 P_{ji}}{\partial b_i \partial c_i} &= D a_i P_{ji}^* Q_{ji}^*, && \text{(A.6)} \\
\frac{\partial^2 P_{ji}}{\partial c_i^2} &= \frac{\partial Q_{ji}^*}{\partial c_i} = 0. && \text{(A.7)}
\end{aligned}$$

Com estas expressões obtemos $\partial^2 P_{ji}/(\partial\boldsymbol{\zeta}_i\partial\boldsymbol{\zeta}_i')$. Sejam

$$\begin{aligned}
\boldsymbol{h}_{ji} &= (P_{ji}^*Q_{ji}^*)^{-1}\left(\frac{\partial P_{ji}}{\partial\boldsymbol{\zeta}_i}\right) = \begin{pmatrix} D(1-c_i)(\theta_j-b_i) \\ -Da_i(1-c_i) \\ \frac{1}{P_{ji}^*} \end{pmatrix}, \\
\boldsymbol{H}_{ji} &= (P_{ji}^*Q_{ji}^*)^{-1}\left(\frac{\partial^2 P_{ji}}{\partial\boldsymbol{\zeta}_i\partial\boldsymbol{\zeta}_i'}\right) \\
&= \begin{pmatrix} D^2(1-c_i)(\theta_j-b_i)^2(1-2P_{ji}^*) & . & . \\ -D(1-c_i)\{1+Da_i(\theta_j-b_i)(1-2P_{ji}^*)\} & D^2a_i^2(1-c_i)(1-2P_{ji}^*) & . \\ -D(\theta_j-b_i) & Da_i & 0 \end{pmatrix}.
\end{aligned}$$

Com isso, de (3.7) temos que

$$\begin{aligned}
\boldsymbol{h}(\boldsymbol{\zeta}_i) &\equiv \frac{\partial \log L(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}_i} \\
&= \sum_{j=1}^{n}\left\{(u_{ji}-P_{ji})\frac{W_{ji}}{P_{ji}^*Q_{ji}^*}\right\}(P_{ji}^*Q_{ji}^*)\boldsymbol{h}_{ji} \\
&= \sum_{j=1}^{n}(u_{ji}-P_{ji})W_{ji}\boldsymbol{h}_{ji}. \qquad \text{(A.8)}
\end{aligned}$$

Retornando a (A.1),

$$\begin{aligned}
\boldsymbol{H}(\boldsymbol{\zeta}_i) &\equiv \frac{\partial \log L(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}_i\partial\boldsymbol{\zeta}_i'} \\
&= \sum_{j=1}^{n}\left\{\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)(P_{ji}^*Q_{ji}^*)\boldsymbol{H}_{ji} - \left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)^2(P_{ji}^*Q_{ji}^*)^2\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\right\} \\
&= \sum_{j=1}^{n}(u_{ji}-P_{ji})W_{ji}\left\{\boldsymbol{H}_{ji}-(u_{ji}-P_{ji})W_{ji}\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\right\}. \qquad \text{(A.9)}
\end{aligned}$$

A.2 Expressões da página 46

Para chegarmos às expressões de $H(\theta_j)$ e $h(\theta_j)$ usadas em (3.40), notemos que de (3.35),

$$\begin{aligned}\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j^2} &= \sum_{i=1}^{I} \left\{ \left[\frac{\partial}{\partial \theta_j} \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) \right] \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) + \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) \left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2} \right) \right\} \\ &= \sum_{i=1}^{I} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) \left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2} \right) - \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right)^2 \left(\frac{\partial P_{ji}}{\partial \theta_j} \right)^2 \right\} \quad \text{(A.10)}\end{aligned}$$

A segunda parcela em (A.10) é obtida por (3.37). Com relação à primeira, temos

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

Sejam

$$h_{ji} = (P_{ji}^* Q_{ji}^*)^{-1} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) = D a_i (1 - c_i), \quad \text{(A.12)}$$

$$H_{ji} = (P_{ji}^* Q_{ji}^*)^{-1} \left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2} \right) = D^2 a_i^2 (1 - c_i)(1 - 2P_{ji}^*). \quad \text{(A.13)}$$

Com isso, de (3.38) temos que

$$\begin{aligned} h(\theta_j) &\equiv \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j} \\ &= \sum_{i=1}^{I} \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^* Q_{ji}^*} \right\} (P_{ji}^* Q_{ji}^*) h_{ji} \\ &= \sum_{i=1}^{I} (u_{ji} - P_{ji}) W_{ji} h_{ji}. \end{aligned}$$

Retornando a (A.10),

$$\begin{aligned} H(\theta_j) &\equiv \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_j^2} \\ &= \sum_{i=1}^{I} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) (P_{ji}^* Q_{ji}^*) H_{ji} - \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right)^2 (P_{ji}^* Q_{ji}^*)^2 h_{ji}^2 \right\} \\ &= \sum_{i=1}^{I} (u_{ji} - P_{ji}) W_{ji} \left\{ H_{ji} - (u_{ji} - P_{ji}) W_{ji} h_{ji}^2 \right\}. \end{aligned} \tag{A.14}$$

A.3 Expressões da página 59

Para chegarmos às expressões de $\boldsymbol{H}_{PI}(\widehat{\boldsymbol{\zeta}})$ e $\boldsymbol{h}_{PI}(\boldsymbol{\zeta}_i)$ usadas em (3.72), comecemos adotando a notação

$$v_{ji} = (u_{ji} - P_i) \frac{W_i}{P_i^* Q_i^*} = \frac{(u_{ji} - P_i)}{P_i Q_i}.$$

Segue, de (3.55) e (3.62), que

$$\frac{\partial P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} = \left[v_{ji} \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right) \right] P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta}). \tag{A.15}$$

Segue de (3.59) que a segunda parcela de (3.71) é obtida por

$$\boldsymbol{h}_{i(j)} \equiv \frac{\partial P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta}) / \partial \boldsymbol{\zeta}_i}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta})} = \int_{I\!R} \left[v_{ji} \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right) \right] g_j^*(\theta) d\theta. \tag{A.16}$$

Com relação à primeira parcela de (3.71), notemos que

$$\begin{aligned} \frac{\partial^2 P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} &= \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left\{ \int_{I\!R} \left[v_{ji} \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right)' \right] P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta}) g(\theta|\boldsymbol{\eta}) d\theta \right\} \\ &= \int_{I\!R} \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left\{ \left[v_{ji} \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right)' \right] P(\boldsymbol{u}_{j.}|\theta, \boldsymbol{\zeta}) \right\} g(\theta|\boldsymbol{\eta}) d\theta. \end{aligned} \tag{A.17}$$

Utilizando (A.15) e o desenvolvimento em (3.24), temos

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\zeta}_i}\left\{\left[v_{ji}\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)'\right]P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})\right\} = \\
&= \frac{\partial}{\partial \boldsymbol{\zeta}_i}\left[v_{ji}\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)'\right]P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta}) + v_{ji}\left(\frac{\partial P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i}\right)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)' \\
&= \left[-v_{ji}^2\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)' + v_{ji}\left(\frac{\partial^2 P_i}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}\right)\right]P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta}) + v_{ji}^2 P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)' \\
&= v_{ji}\left(\frac{\partial^2 P_i}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}\right)P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})
\end{aligned}$$

Segue de (A.17) que

$$\frac{\partial^2 P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} = \int_{I\!R} v_{ji}\left(\frac{\partial^2 P_i}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}\right)P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta.$$

Portanto, a primeira parcela em (3.71) pode ser escrita como

$$\boldsymbol{H}_{ii(j)} \equiv \frac{\partial^2 P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})/(\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i')}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})} = \int_{I\!R} v_{ji}\left(\frac{\partial^2 P_i}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}\right)g_j^*(\theta)d\theta. \quad \text{(A.18)}$$

Por (3.62), para $l \neq i$

$$\begin{aligned}
\frac{\partial^2 P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_l \partial \boldsymbol{\zeta}_i'} &= \frac{\partial}{\partial \boldsymbol{\zeta}_l}\left\{\int_{I\!R}\left[v_{ji}\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)'\right]P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta\right\} \\
&= \int_{I\!R} v_{ji}\left(\frac{\partial P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_l}\right)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)' g(\theta|\boldsymbol{\eta})d\theta \\
&= \int_{I\!R} v_{ji}v_{jl}\left(\frac{\partial P_l}{\partial \boldsymbol{\zeta}_l}\right)\left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i}\right)' P(\boldsymbol{u}_{j.}|\theta,\boldsymbol{\zeta})g(\theta|\boldsymbol{\eta})d\theta
\end{aligned}$$

Portanto, para $l \neq i$, a primeira parcela em (3.71) pode ser escrita como

$$\boldsymbol{H}_{il(j)} \equiv \frac{\partial^2 P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})/(\partial\boldsymbol{\zeta}_l\partial\boldsymbol{\zeta}_i')}{P(\boldsymbol{u}_{j.}|\boldsymbol{\zeta},\boldsymbol{\eta})} = \int_{I\!R} v_{ji}v_{jl}\left(\frac{\partial P_l}{\partial\boldsymbol{\zeta}_l}\right)\left(\frac{\partial P_i}{\partial\boldsymbol{\zeta}_i}\right)' g_j^*(\theta)d\theta \tag{A.19}$$

Podemos agora obter as equações de estimação para $\boldsymbol{\zeta}$. Com as expressões (A.2) a (A.7) obtemos $\partial^2 P_i/(\partial\boldsymbol{\zeta}_i\partial\boldsymbol{\zeta}_i')$, $i = 1,\cdots,I$. Sejam

$$\begin{aligned}
\boldsymbol{h}_i &= (P_i^*Q_i^*)^{-1}\left(\frac{\partial P_i}{\partial\boldsymbol{\zeta}_i}\right) = \begin{pmatrix} D(1-c_i)(\theta-b_i) \\ -Da_i(1-c_i) \\ \frac{1}{P_i^*} \end{pmatrix}, \\
\boldsymbol{H}_{ii} &= (P_i^*Q_i^*)^{-1}\left(\frac{\partial^2 P_i}{\partial\boldsymbol{\zeta}_i\partial\boldsymbol{\zeta}_i'}\right) \\
&= \begin{pmatrix} D^2(1-c_i)(\theta-b_i)^2(1-2P_i^*) & . & . \\ -D(1-c_i)\{1+Da_i(\theta-b_i)(1-2P_i^*)\} & D^2a_i^2(1-c_i)(1-2P_i^*) & . \\ -D(\theta-b_i) & Da_i & 0 \end{pmatrix}
\end{aligned}$$

e, para $i \neq l$,

$$\begin{aligned}
\boldsymbol{H}_{il} &= \boldsymbol{h}_i\boldsymbol{h}_l' = (P_i^*Q_i^*)^{-1}(P_l^*Q_l^*)^{-1}\left(\frac{\partial P_i}{\partial\boldsymbol{\zeta}_i}\right)\left(\frac{\partial P_l}{\partial\boldsymbol{\zeta}_l}\right)' \\
&= \begin{pmatrix} D^2(1-c_i)(1-c_l)(\theta-b_i)(\theta-b_l) & -D^2a_l(1-c_i)(1-c_l)(\theta-b_i) & D(1-c_i)(\theta-b_i)/P_l^* \\ -D^2a_i(1-c_i)(1-c_l)(\theta-b_l) & D^2a_ia_l(1-c_i)(1-c_l) & -Da_i(1-c_i)/P_l^* \\ D(1-c_l)(\theta-b_l)/P_i^* & -Da_l(1-c_l)/P_i^* & [P_i^*P_l^*]^{-1} \end{pmatrix}.
\end{aligned}$$

Retornando a (A.18), temos que a primeira parcela de (3.71) pode ser reescrita como

$$\boldsymbol{H}_{ii(j)} = \int_{I\!R}(u_{ji}-P_i)W_i\boldsymbol{H}_{ii}g_j^*(\theta)d\theta$$

e, para $i \neq l$,

$$\boldsymbol{H}_{il(j)} = \int_{I\!R} (u_{ji} - P_i)(u_{jl} - P_l) W_i W_l \boldsymbol{H}_{il} g_j^*(\theta) d\theta \qquad \text{(A.20)}$$

Com isso, chegamos a

$$\begin{aligned} \boldsymbol{H}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_l) &= \frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_l \partial \boldsymbol{\zeta}_i'} \\ &= \sum_{j=1}^{s} r_j \left\{ \boldsymbol{H}_{il(j)} - \boldsymbol{h}_{i(j)} \boldsymbol{h}_{l(j)}' \right\}. \end{aligned} \qquad \text{(A.21)}$$

Basta, agora, definirmos:

$$\boldsymbol{h}_{PI}(\boldsymbol{\zeta}) = \begin{pmatrix} \boldsymbol{h}(\boldsymbol{\zeta}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{\zeta}_I) \end{pmatrix} \qquad \text{e} \qquad \boldsymbol{H}_{PI}(\boldsymbol{\zeta}) = \begin{pmatrix} \boldsymbol{H}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_1) & \cdots & \boldsymbol{H}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_I) \\ \vdots & \vdots & \vdots \\ \boldsymbol{H}(\boldsymbol{\zeta}_I, \boldsymbol{\zeta}_1) & \cdots & \boldsymbol{H}(\boldsymbol{\zeta}_I, \boldsymbol{\zeta}_I) \end{pmatrix}.$$

Referências Bibliográficas

[1] Andersen, E. B. (1973). Conditional inference in multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, *26*, 31-44.

[2] Andersen E. B. (1980). *Discrete Statistical Models with Social Science Applications*. New York: North-Holand Publishing Company.

[3] Andrade, D. F. (1999). *Comparando o Desempenho de Grupos (Populações) de Respondentes Através da Teoria da Resposta ao Item*. Tese apresentada ao Departamento de Estatística e Matemática Aplicada da UFC para o concurso de professor titular.

[4] Andrade, D. F. e Klein, R. (1999). Métodos estatísticos para avaliação educacional : teoria da resposta ao item. *Boletim da ABE, 43*, 21-28.

[5] Andrade, D. F. e Valle, R. C. (1998). Introdução à teoria da resposta ao item: conceitos e aplicações. *Estudos em Avaliação Educacional*, *18*, 13-32.

[6] Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

[7] Baker,F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.

[8] Batista, J.R. (1999). *Valores Plausíveis para Estimação de Parâmetros Populacionais em Modelos da Teoria da Resposta ao Item*. Dissertação de Mestrado. Belo Horizonte: ICEX/UFMG.

[9] Beaton, A. E. and Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics, 17*, 191-204.

[10] Birnbaum, A. (1957). *Efficient design and use of tests of a mental ability for various decision-making problems*, (Series Report No. 58-16. Project No. 7755-23). USAF School of Aviation Medicine, Texas: Randolph Air Force Base.

[11] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Infering an Examinee's Ability. In F.M. Lord & M.R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA:Addison-Wesley.

[12] Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

[13] Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm. *Psychometrika*, *46*, 433-459.

[14] Bock, R. D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, *35*, 179-197.

[15] Bock, R. D. and Zimowski, M. F. (1997). Multiple Group IRT. In *Handbook of Modern Item Response Theory*. W.J. van der Linder e R.K. Hambleton Eds. New York: Spring-Verlag.

[16] Chow, Y.S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.

[17] Dempster, A. P. , Laird, N. M. and 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.

[18] DeRoos, Y. and Allen-Meares, P. (1998). Applications of rasch analysis: exploring differences in depression between african-american and white children. *Journal of Social Service Research, 23*, 93-107.

[19] Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika, 44*, 169-172.

[20] Fundação Carlos Chagas (1997). *Avaliação das Escolas Estaduais de Ensino Fundamental e Ensino Médio do Rio Grande do Norte, 4v.* São Paulo : Fundação Carlos Chagas.

[21] Fundação Carlos Chagas (1998). *Programa de Aceleração da Aprendizagem: avaliação final, avaliação do material didático e apêndice, 3v.* São Paulo : Fundação Carlos Chagas / Instituto Ayrton Senna.

[22] Genz, A. C. and Malik, A. A. (1980). An adaptive alghorithm for numerical integration over a N-retangular region. *J. Comput. Appl. Math.*, *6*, 295-302.

[23] Goldstein, H. (1994). Recontextualizing mental measurement. *Educational Measurement : Issues and Practice, 13*, 16-43.

[24] Goldstein, H. and Wood, R. (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology, 42*, 139-167.

[25] Granger, C. V., Deutsch, A. and Linn, R. T. (1998). Rasch analysis of the functional independence measure (FIMTM) mastery test. *Arch. Phys. Med. Rehabil.*, *79*, 52-57.

[26] Graybill, F. A. (1969). *Introduction to Matrices with Aplications in Statistics.* Belmont, CA: Wadsworth Publishing Company, Inc.

[27] Gulliksen, H. (1950). *Theory of Mental Tests.* New York : John Wiley and Sons.

[28] Haberman, S. (1975). *Maximum Likelihood Estimates in Exponential Response Models*, (Technical Report) Chicago, IL: University of Chicago.

[29] Haley, D. C. (1952). *Estimation of the dosage mortality relationship when the dose is subject to error*, (Technical Report, 15) Stanford, Calif.: Stanford University, Applied Mathematics and Statistics Laboratory.

[30] Hambleton, R.K. and Cook, L.L. (1997). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement, 14*, 75-96.

[31] Hambleton, R. K. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Academic Publishers.

[32] Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park : Sage Publications.

[33] Hedges, L. V. and Vevea, J. L. (1997). A study of equating in NAEP. *Paper presented at The NAEP Validity Studies Panel*. Palo Alto : American Institutes for Research.

[34] Heitjan, D. F. (1991a). Generalized Norton-Simon models of tumour growth. *Statistics in Medicine*, *10*. 1075-1088.

[35] Heitjan, D. F. (1991b). Nonlinear modeling of serial immunologic data: A case study. *Journal of the American Statistical Association*, *86*. 891-898.

[36] Hildebrand, F. B.(1956). *Introduction to Numerical Analysis*. New-York: McGraw-Hill.

[37] Issac, E. and Keller, H. B. (1966). *Analysis of Numerical Methods*. New York: Wiley & Sons.

[38] Kolen, M. J. and Brennan, R. L. (1995). *Test Equating - Methods and Pratices*. New York: Springer.

[39] Linden, W. J. van der and Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York : Springer-Verlag.

[40] Lord, F. M. (1952). A theory of test scores (No. 7). *Psychometric Monograph*.

[41] 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.

[42] Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, *39*, 247-264.

[43] Lord, F. M. (1975). *Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters*, (Research Bulletin RB-75-33). Princeton, NJ: Educational Testing Service.

[44] Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum Associates, Inc.

[45] Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

[46] Marcoulides, G. A. Ed. (1998). *Modern Methods for Business Research*. Mahwah, NJ: Lawrence Erlbaum Reading, MA: Addison-Wesley.

[47] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

[48] Ministério da Educação e do Desporto (1996). *Sistema Nacional de Avaliação da Educação Básica : SAEB 95 - relatório técnico.* São Paulo / Rio de Janeiro : Fundação Carlos Chagas / Fundação Cesgranrio.

[49] Ministério da Educação e do Desporto (1998). *Sistema Nacional de Avaliação da Educação Básica : SAEB 97 - relatório técnico.* Rio de Janeiro : Fundação Cesgranrio.

[50] Mislevy, R. J. (1986a). Bayes modal estimation in item response models. *Psychometrika*, *51*, 177-195.

[51] Mislevy, R. J. (1986b). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3-31.

[52] Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177-196.

[53] Mislevy, R. J. (1992). *Linking Educational Assessments : concepts, issues, methods and prospects.* Princeton : Educational Testing Service.

[54] Mislevy, R. J. and Stocking, M. L. (1989). A Consumer's Guide to LOGISTIC and BILOG. *Applied Psychological Measurement, 13* 57-75.

[55] Mislevy, R. J. and Bock, R. D. (1990). *BILOG 3 : Item Analysis and Test Scoring with Binary Logistic Models.* Chicago : Scientific Software, Inc.

[56] Muraki, E. (1992). A generalized partial credit model : Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

[57] Muraki, E. and Bock, R. D. (1997). *PARSCALE : IRT Based Test Scoring and Item Analysis for Graded Open-Ended Exercices and Performance Tasks.* Chicago : Scientific Software, Inc.

[58] Nelder, J. A. (1961). The fitting of a generalization of the logistic curve. *Biometrika*, *17* , 89-100.

[59] Nelder, J. A. (1962). An alternative form of a generalized logistic equation. *Biometrics*, *18* , 614-616.

[60] Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, *16* (1), 1-32.

[61] Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psycology*, *3*,1-18.

[62] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley & Sons.

[63] Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen : Danish Institute for Educational Research.

[64] Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika, 1*, 33-49.

[65] Samejima, F. A. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph, 17.*

[66] Secretaria de Estado da Educação de São Paulo (1996). *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP : relatório final dos resultados, 3v.* São Paulo : SEE.

[67] Secretaria de Estado da Educação de São Paulo (1997). *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP : relatório final dos resultados, 4v.* São Paulo : SEE.

[68] Sen, P. K., Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction With Applications*. New York: Chapman & Hall.

[69] Soares, J. F. , Martins, M. I. e Assunção, C. N. B. (1998). Heterogeneidade acadêmica dos alunos admitidos na UFMG e PUC-MG. *Estudos em Avaliação Educacional, 17*, 61-72. São Paulo : Fundação Carlos Chagas.

[70] Stroud, A. H. (1971). *Aproximate Calculation os Multiple Integrals*. New Jersey: Prentice Hall, Englewood Cliffs.

[71] Stroud, A. H. and Secrest, D. (1966). *Gaussian Quadrature Formulas.* Englewood Cliffs, New Jersey : Prentice-Hall.

[72] Swaminathan, H. and Gifford J. A. (1983). Estimation of Parameters in the Three-Parameter Latent Trait Model. In D. Weiss (Ed.),*New Horizonts in Testing*. New York: Academic Press.

[73] Tennant, A., Hillman, M., Fear, J., Pickering, A. and Chamberlin, M. A. (1996). Are we making the most of the stanford health assessment questionnaire? *Brit. J. Rheum., 35*, 574-578.

[74] Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika, 11*, 1-13.

[75] Valle, R. C. (1999). *Teoria da Resposta ao Item.*. Dissertação de Mestrado. São Paulo: IME/USP.

[76] Van Dooren, P. and De Ridder, L. (1976). An adaptive alghorithm for numerical integration over a N-retangular cube. *J. Comput. Appl. Math.*, *2*, 207-217.

[77] Vianna, H. M. (1987). *Testes em Educação*. São Paulo : IBRASA

[78] Wilson, D. T. , Wood, R. , Downs, P. K. and Gibbons, R. (1991). *TESTFACT : Test Scoring, Item Statistics and Item Factor Analysis.* Chicago : Scientific Software, Inc.

[79] Wright, B. D. (1968). *Sample-free test calibration and person measurement.* Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N. J. : Educational Testing Service.

[80] Zimowski, M. F. , Muraki, E. , Mislevy, R. J. and Bock, R.D. (1996). *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items.* Chicago : Scientific Software, Inc.

[81] Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24*, 293-308.