Conteúdo

Capítulo 1

Introdução

O mundo vem passando por diversos problemas provocados pela globalização, exigindo dos países uma rápida adaptação às condições impostas por este processo. A política educacional é um dos pontos de fundamental importância a ser trabalhado para fazer frente aos desafios desse mundo em constante mudança.

O Brasil está empenhado em promover reformas na área educacional que permitam superar o quadro de desvantagem em relação aos índices de escolarização e nível de conhecimento que apresentam os países desenvolvidos. Neste contexto, é necessário buscar novas políticas na área educacional e, em especial, referentes a novas metodologias de avaliação educacional.

O presente trabalho vem ao encontro das reais necessidades do país, propondo-se a apresentar alguns modelos estatísticos baseados na Teoria de Resposta ao Item - TRI, uma teoria que, apesar de ter mais de 50 anos, somente nos últimos 15 anos é que vem sendo aplicada em larga escala nas principais avaliações educacionais em diversos países.

A Teoria de Resposta ao Item propõe modelos que representam a relação entre a probabilidade de uma certa resposta a um item e a *habilidade(s)* ou *traço(s) latente(s)* de um indivíduo [Van der Linden & Hambleton, 1997]. O traço latente é uma característica que não pode ser observada diretamente, e portanto, esse tipo de variável deve ser inferida a partir da observação de variáveis secundárias que estejam relacionadas a ela. Um exemplo de traço latente seria a habilidade ou proficiência (perfeito conhecimento, competência ou capacidade) do indivíduo em uma determinada área de conhecimento. Assim podemos estabelecer relações entre o desempenho num teste (dado pelo escore) e a habilidade. Uma das grandes vantagens da TRI é que ela permite a comparação do desempenho entre populações distintas, desde que submetidas a testes com itens comuns. O mesmo não ocorre com a Teoria Clássica de Medidas, a qual está sujeita à prova como um todo ou a provas paralelas de difícil construção [Vianna, 1987]. Isto porque uma das características mais importantes da TRI é ter como elementos centrais os *itens*. Maiores detalhes sobre os fundamentos e aplicações da TRI ver Lord (1980).

Alguns pontos têm sido levantados na literatura sobre a adequação desta teoria. Dois deles considerados importantes são: a *equalização* ou *comparação* de diferentes avaliações [Andrade, 1999] e a *dimensionalidade do espaço de traços latentes* envolvidos na avaliação [Nojosa, 2001].

Um terceiro ponto, levantado por Mislevy (1991), está relacionado com a qualidade da *estimação da distribuição das habilidades* dos elementos de uma população. O autor discute a possibilidade de se obterem *melhores estimativas da variabilidade*, utilizando-se também outras informações dos respondentes que possam estar associadas com suas habilidades. Exemplos destas informações seriam o grau de escolaridade dos pais, o hábito cultural 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. Algumas das dificuldades para a aplicação desta teoria seriam a inexistência comercial de programas computacionais, a obtenção de informações adicionais fidedignas relevantes ao problema e a inclusão destas informações no modelo.

Um outro ponto que tem sido pouco explorado e que será o principal objetivo deste trabalho é o estudo dos *modelos multivariados*, os quais consideram a possível estrutura de covariância existente entre as habilidades de um indivíduo associadas às diversas disciplinas. Thomas e Gan (1997) descrevem um modelo utilizado pelo NAEP - “National Assesment of Educational Progress” e Tavares (2001) faz uma generalização do modelo de várias populações independentes, introduzido por Bock & Zimowski (1997), para o caso de avaliações de uma mesma população ao longo do tempo, ou seja, para modelos longitudinais.

Na maioria das avaliações educacionais realizadas no Brasil, apesar de terem como objetivo a avaliação do rendimento escolar em mais de uma disciplina do conhecimento, cada respondente é submetido a prova de somente uma das disciplinas. Um exemplo seria o Sistema Nacional de Avaliação do Ensino Básico (SAEB) que avalia o ensino médio no Brasil nas disciplinas Matemática, Português, Biologia, Física, Química, História e Geografia, no qual cada aluno selecionado responde a somente uma prova de uma das disciplinas. A análise de tais provas tem sido realizada a partir da aplicação de modelos univariados a cada uma das disciplinas. Outras avaliações, como por exemplo os exames seletivos para o ensino superior, apesar de submeterem os candidatos a provas de várias disciplinas, também realizam as avaliações através de modelos univariados, não levando em conta a possível estrutura de covariância (ou correlação) existente entre os rendimentos nas disciplinas avaliadas. Mais recentemente, o Instituto Nacional de Estudos e Pesquisas Educacionais do Ministério de Educação e Cultura (INEP/MEC) introduziu a avaliação do Projeto PDE/FUNDESCOLA em 6 estados, incluindo a região metropolitana do Recife, que prevê o acompanhamento de 15.000 alunos da 4^asérie em 1999 por 5 anos nas disciplinas Matemática e Português.

O nosso trabalho terá como principal objetivo o estudo de Modelos Multivariados. Dois principais objetivos serão: obter *melhores estimativas das habilidades* e *estimativas* das possíveis *correlações* entre as *habilidades* em *diferentes disciplinas*. Com este intuito, o Capítulo 2 trata dos principais modelos univariados da TRI, com suas interpretações, suposições básicas e os principais métodos de estimação dos parâmetros dos itens e habilidades dos respondentes. O Capítulo 3 considera uma proposta de modelo multivariado, juntamente com o método de estimação para os parâmetros dos itens. No Capítulo 4 considera-se a estimação dos parâmetros populacionais seguido pelo Capítulo 5 em que apresentamos a estimação conjunta dos parâmetros. Alguns procedimentos bayesianos são apresentados no Capítulo 6. No Capítulo 7 apresentamos aplicações a dados simulados e a dados reais e algumas considerações serão apresentadas no Capítulo 8. No Apêndice descrevemos alguns resultados importantes para o desenvolvimento do trabalho.

Capítulo 2

Modelos Matemáticos Univariados

Os modelos da TRI propõem uma relação entre a(s) habilidade(s) de um indivíduo e a probabilidade de uma certa resposta a um determinado item; de forma que, quanto maior a(s) habilidade(s) do indivíduo maior será a probabilidade de acerto. Em geral, os modelos propostos na literatura dependem fundamentalmente de três fatores:

(i) da natureza do item - dicotômico ou não-dicotômico;

(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 - unidimensional ou multidimensional, respectivamente.

É importante discutirmos aqui a sutil diferença existente entre os *modelos multivariados* e os *modelos multidimensionais*, pois, apesar de nos referirmos a duas ou mais quantidades de traços latentes, os modelos multivariados referem-se a traços distintos, os quais são avaliados por diferentes testes. De forma similar, os modelos multidimensionais consideram vários traços latentes; no entanto, estes traços explicam uma única habilidade (ou traço latente); ou seja; a habilidade é explicada por um vetor de habilidades, sendo necessária para sua avaliação a aplicação de um único teste com características multidimensionais.

2.1 Modelos envolvendo um único grupo

Para melhor compreensão dos modelos é importante ter em mente o conceito de grupo e de população. *Grupo* é o conjunto de indivíduos que compõem a amostra de uma população. Portanto, o conceito de grupo está diretamente associado ao processo de amostragem e estaremos sempre considerando o processo de amostragem aleatória simples.

População é definida por determinadas características que podem variar de acordo com os objetivos do estudo. Estas características podem ser ou não relevantes para diferenciar uma população de outra, dependendo do caso. Exemplos usuais de populações distintas são: séries distintas (3ª série e 4ª série do Ensino Fundamental); períodos distintos (diurno e noturno); uma mesma série, mas em anos distintos (3ª série de 1999 e 3ª série de 2000), etc.

2.1.1 Modelos para itens dicotômicos ou dicotomizados

Apenas o *modelo logístico unidimensional de 3 parâmetros* (*ML3*), o qual pertence aos *Modelos Logísticos* para itens dicotômicos ou dicotomizados (itens com mais de 2 categorias de resposta ou de resposta aberta, mas corrigidos como certo ou errado) será tratado aqui. Na prática, este modelo tem sido o mais utilizado e, além disso, os *modelos logístico de 2 e 1 parâmetro* (*ML2* e *ML1*) podem ser facilmente obtidos através de algumas restrições dos parâmetros do ML3. Este modelo é dado por

$$P(U_{ji}=1|\theta_j)=c_i+(1-c_i)\frac{1}{1+e^{-Da_i(\theta_j-b_i)}},\ j=1,2,...,N,\ \text{e}\ i=1,2,...,n, \tag{2.1}$$

onde

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

θ_j representa a habilidade ou traço latente do j-ésimo indivíduo,

$P(U_{ji} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente ao item i,

a_i é o parâmetro de discriminação (ou de inclinação) do item i,

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

c_i é o parâmetro que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente ao item i (muitas vezes referido como a probabilidade de resposta correta dada ao acaso), e

D é um fator de escala, constante e igual a 1. Utiliza-se $D = 1,7$, quando se deseja aproximar os resultados da função logística aos da função ogiva normal, o qual foi o primeiro modelo da TRI que argumenta a função de distribuição acumulada da normal como a função de resposta ao item ou *curva característica do item* [Van der Linden & Hambleton, 1997].

Interpretação do Modelo

Como já comentado anteriormente sobre os modelos da TRI, observa–se na Figura 2.1 que quanto maior a habilidade, maior é a probabilidade de uma resposta correta ao item. Este gráfico apresenta a Curva Característica do Item - CCI, a qual tem a forma de "S" com inclinação e deslocamento na escala de habilidade definidos pelos parâmetros do item.

As Figuras 2.2 e 2.3 representam curvas características dos itens com diferentes valores para os parâmetros a, b e c. Note na Figura 2.2 que, para um fixado valor de a, a habilidade requerida para uma resposta correta a um item com uma mesma probabilidade aumenta à medida em que o valor de b também aumenta, o que justifica a denominação do parâmetro b de parâmetro de dificuldade (ou de posição) e que seu valor esteja na mesma escala da habilidade. Por outro lado, na Figura 2.3 observa-se que o valor do parâmetro a é proporcional ao coeficiente angular da reta tangente no ponto de inflexão da CCI, pois quanto maior o valor de a, mais acentuada é a inclinação da curva, tendo como conseqüência uma maior diferença entre as probabilidades de resposta correta de dois indivíduos com habilidades distintas, mesmo que ligeiramente.

Figura 2.1: Exemplo de uma Curva Característica de Item - CCI

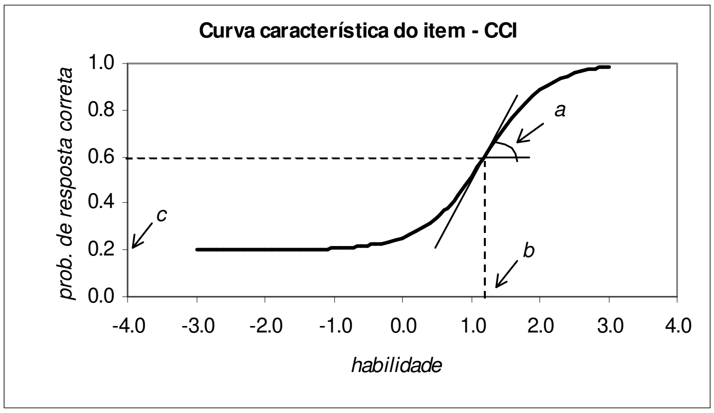
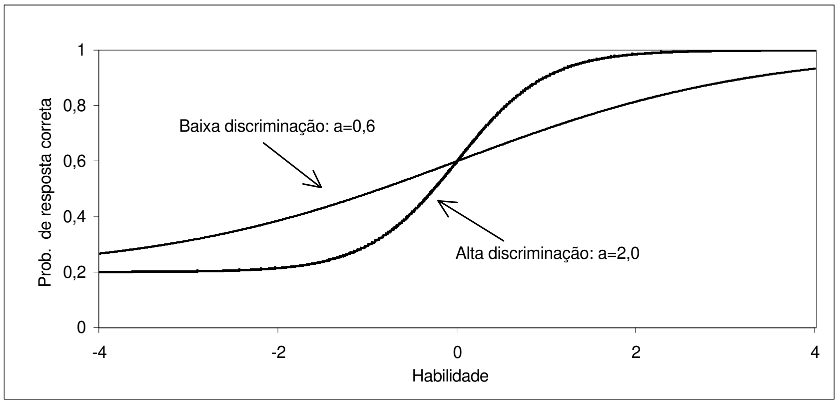


Figura 2.2: Exemplo de itens com grau de dificuldade baixo, médio e alto.
a=2,0 b=-2,0(2)2,0 c=0,20

Figura 2.3: Exemplo de itens com grau de discriminação baixo e alto.
a=0,6, 2,0 b=0,0 c=0,20



O parâmetro c nada mais é do que a probabilidade de resposta correta de um indivíduo com baixa habilidade e, portanto, não depende de escala. É também denominado de parâmetro de acerto casual.

Teoricamente os parâmetros de *dificuldade* b e de *discriminação* a podem assumir qualquer valor real entre $-\infty$ e $+\infty$. Sendo evidente que nesse modelo não se espera um valor negativo para o parâmetro a, pois se o fosse teríamos a condição de que a probabilidade de responder corretamente o item diminuiu com o aumento da habilidade. É importante observarmos que estes parâmetros (b e a), diferentemente da Teoria Clássica [Vianna, 1987], não medem proporção e correlação, respectivamente.

2.1.2 Outros modelos

Detalhes dos diversos modelos existentes podem ser vistos em Van der Linden & Hambleton (1997) e Andrade, Tavares & Valle (2000). Alguns modelos apresentados por estes autores, são:

Modelo Normal de 3 *parâmetros*, o qual foi o primeiro proposto na TRI.

Modelo Logístico de 2 *parâmetros*, obtido ao se supor que não há possibilidade de resposta ao acaso ($c_i = 0$) no *ML3*; se além desta suposição, supusermos que o poder de discriminação é o mesmo para todos os itens envolvidos em um teste, obtemos o *Modelo*

Logístico de 1 *parâmetro* ou *Modelo Rasch.*

Outros modelos têm surgido admitindo-se o formato de respostas politômicas, abertas, o tempo necessário para se realizar um teste ou tarefa e modelos que consideram a existência de múltiplas habilidades ou componentes cognitivos.

2.2 Modelos envolvendo dois ou mais grupos

Extensões dos modelos citados anteriormente já foram desenvolvidas para um teste que envolve mais de uma população. No entanto, a generalização dos modelos logísticos unidimensionais de 1, 2 e 3 parâmetros recentemente proposta por Bock & Zimowski (1997) é a que vem sendo utilizada na prática, por ser um dos poucos modelos já implementados computacionalmente e que tem fornecido resultados muito bons. Este modelo é dado por

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

com $j = 1, 2, ..., N_k$, $i = 1, 2, ..., n$ e $k = 1, 2, ..., K$, onde

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

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

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

(a_i, b_i, c_i) são os mesmos parâmetros de item descritos anteriormente.

Nesse modelo considera-se a existência de K populações independentes, e como existem itens comuns, a estimação dos parâmetros é feita conjuntamente. Uma distribuição é considerada para cada população K, cuja média é μ_k e variância σ_k^2, $k = 1, ..., K$. Em geral, é adotado $\theta_k \sim N(\mu_k, \sigma_k^2)$ sendo necessário adotar-se uma população de referência, que por simplicidade é dada por uma *distribuição normal padrão*. Esse modelo representa um dos principais avanços na Teoria da Resposta ao Item nos últimos anos, por permitir que comparações de indivíduos de grupos distintos submetidos a provas diferentes, mas com itens comuns, seja feita diretamente no processo de estimação, evitando assim erros associados a cada método de comparação utilizado anteriormente. Em Andrade (1999) encontra-se comparações de vários métodos de comparação de populações (*equalização*).

2.3 Estimação

Na TRI a probabilidade de uma resposta correta num determinado item depende da habilidade do indivíduo e dos parâmetros que caracterizam o item. Na prática, o problema mais realista surge quando ambos são desconhecidos e o que é conhecido são as respostas dos indivíduos aos itens do teste. O processo de estimação dos parâmetros do item é chamado de calibração [Baker, 1992] e atualmente, como já existem o que se denomina de banco de itens, ou seja, um banco de dados com itens já calibrados, é possível também encontrar situações em que se deseja estimar apenas as habilidades dos indivíduos. O mesmo pode ocorrer se tivermos as habilidades estimadas em um teste e desejamos utilizá-las para a calibração de um novo conjunto de itens.

Para descrever o processo de estimação, utilizaremos a seguinte notação:

$\mathbf{U}_{j.} = (U_{j1}, U_{j2}, ..., U_{jn})'$ o vetor de respostas aos n itens do indivíduo j,

$\mathbf{U}_{..} = (\mathbf{U}'_{1.}, \mathbf{U}'_{2.}, ..., \mathbf{U}'_{N.})'$ o vetor total de respostas.

Então, usando as suposições de *Unidimensionalidade* e de *Independência Local* ou *Independência Condicional*, a qual diz que, fixada a habilidade θ de um indivíduo, as respostas aos itens são independentes, chegamos à seguinte função de verossimilhança:

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{U}_{..}) = \prod_{j=1}^{N} P(\mathbf{U}_{j.}|\theta_j) = \prod_{j=1}^{N}\prod_{i=1}^{n} P(\theta_j; \boldsymbol{\beta}_i)^{u_{ji}} \left[1 - P(\theta_j; \boldsymbol{\beta}_i)\right]^{1-u_{ji}},$$

onde $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_N)'$ é o vetor de habilidades, $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, ..., \boldsymbol{\beta}'_n)'$ é o vetor cujos elementos $\boldsymbol{\beta}_i = (a_i,\ b_i,\ c_i)'$ são os vetores de parâmetros do item i e $P(\theta_j; \boldsymbol{\beta}_i)$ é a probabilidade de uma certa resposta ao item i pelo $j - \acute{e}simo$ indivíduo.

Os valores de $\boldsymbol{\theta}$ e $\boldsymbol{\beta}$ que maximizam a função de verossimilhança são os estimadores de máxima verossimilhança (EMV). Estes valores são equivalentes aos que maximizam o logaritmo da função de verossimilnahça (FV), de forma que após derivar e igualar as primeiras derivadas a zero, temos como resultado o seguinte sistema de equações:

$$
\begin{aligned}
a_i &: \ D\sum_{j=1}^{N}(u_{ji} - P_{ji})\,(\theta_j - b_i)\,\frac{P_{ji}^*}{P_{ji}} = 0, \qquad i = 1, ..., n, \\
b_i &: \ -Da_i\sum_{j=1}^{N}(u_{ji} - P_{ji})\frac{P_{ji}^*}{P_{ji}} = 0, \qquad i = 1, ..., n, \\
c_i &: \ \sum_{j=1}^{N}(u_{ji} - P_{ji})\frac{1}{(P_{ji} - c_i)}\frac{P_{ji}^*}{P_{ji}} = 0, \qquad i = 1, ..., n, \\
\theta_j &: \ D\sum_{i=1}^{n}(u_{ji} - P_{ji})a_i\frac{P_{ji}^*}{P_{ji}} = 0, \qquad j = 1, ..., N,
\end{aligned}
$$

onde $P_{ji} \equiv P(\theta_j; \boldsymbol{\beta}_i)$ e $P_{ji}^* = \left\{1 + e^{-Da_i(\theta_j - b_i)}\right\}^{-1}$ é o modelo logístico de 2 parâmetros.

Em geral, a estimação é feita por algum método iterativo, como *Newton-Raphson* ou *Escore* de *Fisher*; podendo ser abordado conjuntamente, sendo mais conhecido por método de *Máxima Verossimilhança Conjunta - MVC*; ou em duas etapas, conhecido por método de *Máxima Verossimilhança Marginal - MVM.*

Na abordagem de *MVC*, o número de parâmetros pode ser extremamente grande $(3n + N)$, o que exige um grande esforço computacional. Birbaum em 1968 propôs um processo *vai e volta*, o qual inicia-se com valores iniciais para as habilidades (escores padronizados, por exemplo), resolve-se as equações em função dos parâmetros dos itens, e prossegue com a estimação das habilidades ao considerar os parâmetros dos itens estimados anteriormente como conhecidos; e assim sucessivamente até que algum critério de parada do processo seja satisfeito. Problemas de convergência surgem neste procedimento, pois o número de parâmetros θ_j, $j = 1, ..., N$, cresce com o número de indivíduos; os quais, na estimação conjunta são denominados de *parâmetros incidentais*, enquanto que o número dos parâmetros dos itens, conhecidos como *parâmetros estruturais*, não se alteram quando a amostra cresce (denominações devidas a Neyman & Scott, 1948). Por este motivo a estimação em duas etapas, feita pelo método de *Máxima Verossimilhança Marginal - MVM* tornou-se popular para os modelos de 2 e 3 parâmetros.

O método de *MVM* foi introduzido por Bock & Lieberman (1970) para o modelo da ogiva normal, mas foi imediatamente adaptado para o modelo logístico. Este método considera uma função densidade de probabilidade (fdp) $g(\theta|\boldsymbol{\eta})$ que descreve a distribuição da habilidade populacional, onde $\boldsymbol{\eta}$ é o conjunto de parâmetros associados à população. A probabilidade marginal obtida para um vetor resposta $\mathbf{U}_{j.} = (U_{j1}, U_{j2}, ..., U_{jn})'$ sobre os itens num teste é igual a

$$P(\mathbf{U}_{j.}|\boldsymbol{\beta}, \boldsymbol{\eta}) = \int_{-\infty}^{\infty} \prod_{i=1}^{n} P(\theta; \boldsymbol{\beta}_i)^{u_{ji}} \left[1 - P(\theta; \boldsymbol{\beta}_i)\right]^{1-u_{ji}} g(\theta|\boldsymbol{\eta})\, d\theta. \tag{2.2}$$

Desta forma, a função de verossimilhança marginal associada ao vetor de dados completo $\mathbf{U}_{..}$ pode ser expressa pelo que se denomina de Multinomial de Kernel, dada por

$$L(\boldsymbol{\beta}, \boldsymbol{\eta}; \mathbf{U}_{..}) = \prod_{j=1}^{2^n} \pi_j^{r_j},$$

onde π_j é a probabilidade do padrão de resposta dada em (2.2), a qual tem frequência r_j na matriz de dados. Após estimados os parâmetros dos itens sem a influência da habilidade, pode-se estimar as habilidades individualmente por MV, considerando-se os parâmetros dos itens conhecidos ou, se desejado, pela moda ou média da distribuição a posteriori de θ dada por

$$g_j^*(\theta) = P(\theta|\mathbf{U}_{j.}, \boldsymbol{\beta}, \boldsymbol{\eta}) = \frac{P(\mathbf{U}_{j.}|\theta, \boldsymbol{\beta}) g(\theta|\boldsymbol{\eta})}{P(\mathbf{U}_{j.}|\boldsymbol{\beta}, \boldsymbol{\eta})}, j = 1, 2, ..., N.$$

Devido ao grande número de padrões de respostas possíveis (2^n), o método de Bock &

Lieberman é lento e impraticável para mais de 10 itens. Entretanto, Bock e Aitkin (1981) introduziram uma nova versão deste método. Nesta versão foi adicionada a suposição de *independência entre os itens*, de forma que seja possível estimar os itens individualmente; além disso foi adaptada a aplicação do algoritmo EM de Dempster, Laird & Rubin (1977). É importante notar que as suposições de independência local e independência entre os itens são completamente diferentes, uma vez que a primeira é inerente às respostas dos indivíduos, enquanto a segunda se refere apenas aos itens.

Uma variedade de aproximações bayesianas para a estimação dos parâmetros do *ML2* e *ML3* tem sido estudada. Referências relevantes são Mislevy (1986), Swaminathan & Gifford (1985, 1986) e Tsutakawa & Lin (1986).

2.4 A Escala da Habilidade

Verificamos anteriormente que o parâmetro de dificuldade b está definido na mesma escala da habilidade e portanto pode ser considerado como uma habilidade. Na verdade, é a habilidade necessária para uma probabilidade de acerto igual a $(c+1)/2$ no ML3. Esta relação é de fundamental importância para a interpretação dos resultados obtidos após o processo de estimação, pois é nesta fase que especialistas da área pedagógica, psicológica, entre outros da área educacional, contribuem ao discutir de forma crítica sobre a validade pedagógica dos resultados obtidos com a aplicação da TRI.

Um problema de escala, conhecido por *não-identificabilidade* ou *indeterminação*, surge quando desejamos estimar tanto as habilidades quanto os parâmetros dos itens, ou seja, quando desconhecemos todos os parâmetros. Neste caso, as habilidades e os parâmetros dos itens não são unicamente determinados. Se substituirmos θ por $\theta^* = \alpha\theta + \beta$, b por $b^* = \alpha b + \beta$ e a por $a^* = a/\alpha$, $\alpha > 0$ e β real em (2.1), verificaremos que a probabilidade de resposta correta ao item permanece a mesma, ou seja, $P(\theta; \boldsymbol{\beta}_i) = P(\theta^*; \boldsymbol{\beta}_i^*)$ e dado que α e β são constantes arbitrárias, a função de verossimilhança não terá um único ponto de máximo.

Desta forma, qualquer método aplicado para encontrar o máximo da função verossimilhança falhará devido a esta não-identificabilidade. Por este motivo, é necessário estabelecer uma escala de medida, o que pode ser feito escolhendo-se um ponto de referência e uma medida de dispersão. Isto ocorre sempre que tratamos de variáveis latentes, como por exemplo na análise fatorial. No processo iterativo de estimação, deve-se padronizar as estimativas das habilidades de forma que estas tenham sempre média μ e variância σ^2; em geral, é adotado $\mu = 0$ e $\sigma^2 = 1$, denotado por $(0;1)$ ou $(0,1)$. No caso da estimação de MVM, freqüentemente supõe-se $\theta \sim N(0,1)$. Vale lembrar que esta não é a única forma de se solucionar a não-identificabilidade (ver Baker, (1992)).

Quando o interesse é a estimação dos parâmetros em modelos envolvendo mais de uma população, o problema de indeterminação também existe, mas neste caso não é suficiente estabelecer escalas de medida ou distribuições a priori para as habilidades, como no caso de um único grupo. Pois, devido a existência de mais de um grupo de respondentes, é necessário também que se defina uma das populações como a *população de referência*. Desta forma, todos os parâmetros populacionais serão medidos na mesma escala (métrica) da população de referência.

A necessidade de se ter uma população de referência surge de um problema de bastante interesse prático, o qual é denominado de *Equalização* ("*Equating*") ou *comparação de vários grupos*. Apesar dos parâmetros serem estimados numa mesma escala $(0,1)$, estas não seriam comparáveis, porque cada escala estaria vinculada a uma população. Isto quer dizer, por exemplo, que uma habilidade 0 numa população pode representar ter mais (ou menos) conhecimento numa determinda área, do que uma outra população com a mesma habilidade 0. O mesmo ocorre com o parâmetro b. Por exemplo, um parâmetro de dificuldade b com valor negativo numa população poderá ter seu valor estimado numa segunda com um valor positivo, não implicando que este parâmetro teve seu valor modificado, mas sim que o item é considerado fácil numa população, enquanto é relativamente difícil para outra. Esta importante propriedade da TRI é conhecida por *Princípio da Invariância*. Este princípio diz que as estimativas das habilidades não se alteram mediante a apresentação de um novo conjunto de itens e, por outro lado, as estimativas dos parâmetros dos itens não se alteram ao se aplicar os mesmos itens a populações distintas [Hambleton et al, 1991]. Em suma, para que seja possível fazer *comparações entre populações* ou *equalização*, é necessário colocar tais parâmetros na mesma escala.

Vários são os métodos de equalização, podendo ser implementados *via população* ou *via itens comuns*. No primeiro usa-se o fato de que um único grupo de respondentes é aleatoriamente submetido à provas distintas, bastando que todos os itens sejam calibrados conjuntamente para garantir que todos estejam na mesma métrica. Já na equalização via itens comuns, esta garantia é dada pelos itens comuns entre as populações, servindo de ligação entre elas. Os principais métodos são: Método de Regressão, Média-Desvio ("Mean-Sigma"), Média-Desvio Robusto e Curva Característica. Para mais detalhes sobre o assunto sugerimos a leitura de Andrade (1999) e Valle (1999).

Capítulo 3

Modelo Multivariado para uma Única População: Estimação dos Parâmetros dos Itens

Neste capítulo será apresentada uma proposta de Modelo Multivariado, a qual teve como origem os Modelos Longitudinais de Tavares (2001). Tavares propôs uma generalização do Modelo de Bock & Zimowski (1997) [sec. 2.2] ao considerar a situação em que os dados possuem uma estrutura longitudinal. Desta forma, é possível, por exemplo, analisar o desempenho educacional de indivíduos ao longo de T condições de avaliação (tempo). A nossa proposta consiste em contruir um modelo em que uma única população é avaliada em relação a T habilidades ou traços latentes, levando-se em consideração uma possível estrutura de correlação entre as habilidades, como foi considerado nos modelos longitudinais de Tavares.

Após construir o modelo, apresentaremos o primeiro bloco de estimação dos parâmetros. Este bloco consiste na estimação dos parâmetros dos itens considerando-se os parâmetros populacionais fixos e conhecidos. Deixaremos para o próximo capítulo a apresentação do segundo bloco de estimação, em que temos o caso contrário, pois a estimação dos parâmetros populacionais é feita ao se considerar os parâmetros dos itens fixos e conhecidos. Lembramos que atualmente tais situações são possíveis, especialmente em relação aos parâmetros dos itens, os quais são comumente armazenados em *bancos de itens*. Finalmente, no capítulo 5 apresentamos a situação mais realista em que desconhecemos tanto os parâmetros dos itens quanto os populacionais e, portanto, devemos fazer a junção dos dois blocos para realizar a estimação conjunta dos parâmetros.

3.1 Construção do Modelo Multivariado

Neste modelo consideramos a existência de uma única população sobre a qual desejamos estimar a proficiência em K distintas áreas do conhecimento $\boldsymbol{\theta} = (\theta_1, ..., \theta_K)'$. Iremos supor que os indivíduos selecionados numa amostra, ou o grupo de indivíduos, são provenientes de uma amostra aleatória simples.

É importante ressaltarmos uma *particularidade* deste modelo em relação aos modelos longitudinais, pois o modelo que propomos *não apresenta itens comuns* entre os *diversos testes*, enquanto que nos estudos longitudinais a existência de itens comuns é de fundamental importância para se realizar a equalização. Desta maneira, cada teste ou prova apresentará n_k itens e, portanto, um total de $n = \sum_{k=1}^{K} n_k$ itens será apresentado a cada indivíduo.

3.1.1 Modelo Probabilístico

Nesta modelagem, utilizaremos o modelo logístico de 3 parâmetros para a probabilidade de resposta correta a um item, pois este é um modelo de grande aceitação entre os profissionais das áreas de avaliação. No entanto, isto não impede a utilização de outros modelos.

Se U_{jik} é a resposta (binária) do $j-ésimo$ indivíduo ao $i-ésimo$ item do $k-ésimo$ teste (prova, disciplina ou avaliação), então

$$P(U_{jik} = 1|\theta_{jk}; \boldsymbol{\beta}_{ik}) = c_{ik} + (1 - c_{ik})\{1 + \exp[-Da_{ik}(\theta_{jk} - b_{ik})\}^{-1}, \tag{3.1}$$

onde θ_{jk} é a habilidade do indivíduo j no teste k, $\boldsymbol{\beta}_{ik} = (a_{ik}, b_{ik}, c_{ik})$ é o vetor de parâmetros associado ao item i no teste k e assumiremos $D = 1$.

Algumas suposições são necessárias, como, por exemplo, a suposição de que a probabilidade de acerto a qualquer item, em qualquer dos testes, pode ser modelada pela mesma função resposta (ML3). Além disso, assumem-se as seguintes suposições de independência:

(a) Independência Local ou Condicional. Já citada anteriormente, esta suposição diz que uma vez conhecida a habilidade θ_{jk} do indivíduo, as *respostas aos itens* que compõem cada teste são independentes;

(b) Uma extensão da independência local declara que as *respostas aos itens de um teste* são fornecidos de forma independente dos outros testes, uma vez conhecida e fixada as k habilidades do indivíduo, $\boldsymbol{\theta}_j = (\theta_{j1}, ..., \theta_{jK})^{'}$;

(c) Respostas ao teste k dependem apenas da habilidade θ_{jk}, ou seja, independem da habilidade θ_{jl} uma vez conhecida θ_{jk}, $k \neq l$;

(d) Indivíduos diferentes respondem aos testes de forma independente.

Outras suposições são freqüentes, como: (i) a de que o tempo destinado ao teste é suficiente para que todos possam responder a todos os itens, (ii) a ordem em que os itens são apresentados aos indivíduos não interfere no desempenho dos mesmos e, principalmente, (iii) que a probabilidade de acerto a cada item é função de um único traço latente, ou seja, a suposição de unidimensionalidade.

Em relação ao vetor de habilidades $\boldsymbol{\theta}$ da população, consideraremos uma distribuição contínua multivariada com vetor de parâmetros $\boldsymbol{\eta}$ de componentes finitas, cuja fdp será denotada por $g(\boldsymbol{\theta}|\boldsymbol{\eta})$. No caso Normal $K - variado$ temos que $\boldsymbol{\eta} = (\boldsymbol{\mu}, \Sigma)$, onde $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_K)^{'}$ é o vetor de médias e $\Sigma = (\sigma_{kl})_{k,l \leq K}$ é a matriz de covariâncias.

3.1.2 Construção da Verossimilhança

O método utilizado para a estimação dos parâmetros dos itens $\boldsymbol{\beta}$ e dos parâmetros populacionais $\boldsymbol{\eta}$ será basicamente o de máxima verossimilhança. Por isso, apresentaremos a seguir a função de verossimilhança e deixaremos para a próxima seção o desenvolvimento para se chegar às equações de verossimilhança para os parâmetros dos itens.

Consideremos inicialmente a seguinte notação:

$\mathbf{U}_{j.k} = (U_{j1k}, U_{j2k}, ..., U_{jn_kk})'$ o vetor de respostas do indivíduo j no teste k;

$\mathbf{U}_{j..} = (\mathbf{U}'_{j.1}, \mathbf{U}'_{j.2}, ..., \mathbf{U}'_{j.K})'$ o vetor de respostas do indivíduo j em todos os K testes;

$\mathbf{U}_{...} = (\mathbf{U}'_{1..}, \mathbf{U}'_{2..}, ..., \mathbf{U}'_{N..})'$ o vetor total de respostas.

Usando a independência local, podemos escrever a probabilidade condicional associada ao vetor de respostas $\mathbf{U}_{j.k}$ como

$$P(\mathbf{U}_{j.k}|\theta_k, \boldsymbol{\beta}_{.k}) = \prod_{i \in I_k} P(U_{jik}|\theta_k, \beta_{ik}),$$

onde I_k, $k = 1, ..., K$, é o conjunto dos índices dos itens pertencentes ao teste k e $\boldsymbol{\beta}_{.k}$ corresponde ao vetor de parâmetros dos itens relacionados ao teste k.

Utilizando a extensão da indepedência local podemos escrever a probabilidade condicional associada ao vetor de respostas do indivíduo j em todos os testes como

$$P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \prod_{k=1}^{K} P(\mathbf{U}_{j.k}|\theta_k, \boldsymbol{\beta}_{.k}) = \prod_{k=1}^{K} \prod_{i \epsilon I_k} P(U_{jik}|\theta_k, \beta_{ik}). \tag{3.2}$$

Note que a probabilidade marginal do vetor de resposta $\mathbf{U}_{j..}$ é dada por

$$\begin{aligned} P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}) &= \int_{\mathbb{R}^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\beta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta}, \end{aligned} \tag{3.3}$$

a qual é função apenas de $\boldsymbol{\beta}$ e $\boldsymbol{\eta}$, dependendo de $\boldsymbol{\theta}$ somente através de sua distribuição. Assim, utilizando (3.3), podemos escrever a verossimilhança $L(\boldsymbol{\beta}, \boldsymbol{\eta}) = P(\mathbf{U}_{...}; \boldsymbol{\beta}, \boldsymbol{\eta})$ como

$$L(\boldsymbol{\beta}, \boldsymbol{\eta}) = \prod_{j=1}^{N} P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}).$$

Uma abordagem que tem sido bastante utilizada quando o número de indivíduos é grande em relação ao número de itens e que pode implicar em vantagens computacionais é a de *padrão de resposta*.

Desta maneira, como temos n_k itens no teste k, totalizando $n = \sum_{k=1}^{K} n_k$ respostas para cada indivíduo; em itens com 2 possíveis respostas (0 ou 1), há um total de $S = 2^n$ possíveis respostas ou *padrões de resposta*.

Poderíamos assim trabalhar com o índice j representando o padrão de resposta ao invés de indivíduo. Então, se r_j representa o número de ocorrências do padrão de resposta j, $s \leq \min(N, S)$ o número de padrões de resposta com $r_j > 0$, e notando que $\sum_{j=1}^{s} r_j = N$, temos que, usando a independência entre as respostas dos diferentes indivíduos, o vetor de freqüências dos padrões de respostas $\mathbf{R} = (r_1, r_2, ..., r_s)$ segue uma distribuição *Multinomial*, ou seja,

$$P(\mathbf{R}|\boldsymbol{\beta}, \boldsymbol{\eta}) = \frac{N!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} [P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})]^{r_j},$$

o que faz com que a verossimilhança seja dada por

$$L(\boldsymbol{\beta}, \boldsymbol{\eta}) = \frac{N!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} \left\{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})\right\}^{r_j}. \tag{3.4}$$

Note que, se $r_j = 1$, obtemos o caso em que o número de padrões de respostas s é igual ao número total de indivíduos N.

3.2 Equações de Estimação: Caso Geral

Aplicando a função logarítmica em (3.4), obtemos a log-verossimilhança dada por

$$\log L(\boldsymbol{\beta}, \boldsymbol{\eta}) = C + \sum_{j=1}^{s} r_j \log P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}),$$

onde C é a constante proporcional à verossimilhança. Assim as equações de estimação para os parâmetros dos itens são dadas por

$$\frac{\partial \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \beta_{ik}} = \mathbf{0};\ i \in I_k, \quad k = 1, ..., K.$$

Ou seja,

$$
\begin{aligned}
\frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}} &= \frac{\partial}{\partial \beta_{ik}}\left\{\sum_{j=1}^{s} r_j \log P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})\right\} \\
&= \sum_{j=1}^{s} r_j \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}}. \qquad (3.5)
\end{aligned}
$$

Agora notemos que

$$
\begin{aligned}
\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}} &= \frac{\partial}{\partial \boldsymbol{\beta}_{ik}}\left\{\int_{\mathbb{R}^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}\right\} \qquad (3.6) \\
&= \int_{\mathbb{R}^K}\left\{\frac{\partial}{\partial \boldsymbol{\beta}_{ik}} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})\right\} g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&= \int_{\mathbb{R}^K}\left\{\frac{\partial}{\partial \boldsymbol{\beta}_{ik}} \log P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})\right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \qquad (3.7) \\
&= \int_{\mathbb{R}^K}\left\{\frac{\partial}{\partial \boldsymbol{\beta}_{ik}} \sum_{k=1}^{K}\sum_{i \epsilon I_k} \log P(U_{jik}|\theta_k,\boldsymbol{\beta}_{ik})\right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&= \int_{\mathbb{R}^K}\left\{\frac{\partial P(U_{jik}|\theta_k,\boldsymbol{\beta}_{ik})/\partial \boldsymbol{\beta}_{ik}}{P(U_{jik}|\theta_k,\boldsymbol{\beta}_{ik})}\right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}, \qquad (3.8)
\end{aligned}
$$

onde a ordem da derivada e da integral em (3.6) pôde ser permutada com base no Teorema da Convergência Dominada de Lebesgue (Chow & Teicher, 1978).

Agora, consideremos a seguinte notação

$$
\begin{aligned}
P(U_{jik}|\theta_k,\boldsymbol{\beta}_{ik}) &= P(U_{jik}=1|\theta_k,\boldsymbol{\beta}_{ik})^{U_{jik}}\left(1-P(U_{jik}=1|\theta_k,\boldsymbol{\beta}_{ik})\right)^{1-U_{jik}} \\
&= P_{ik}^{U_{jik}} Q_{ik}^{1-U_{jik}}, \qquad (3.9)
\end{aligned}
$$

onde na última igualdade o índice j da probabilidade de sucesso P e de fracasso Q desaparece por $P(U_{jik}=1|\theta_k,\boldsymbol{\beta}_{ik})$ só depender do $j-\acute{e}simo$ padrão de resposta através de U_{jik}.

Usando a notação (3.9) podemos obter

$$\begin{aligned}
\frac{\partial P(U_{jik}|\theta_k,\boldsymbol{\beta}_{ik})}{\partial \boldsymbol{\beta}_{ik}} &= \frac{\partial}{\partial \boldsymbol{\beta}_{ik}}\left(P_{ik}^{U_{jik}}Q_{ik}^{1-U_{jik}}\right) \\
&= U_{jik}P_{ik}^{U_{jik}-1}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)Q_{ik}^{1-U_{jik}} + P_{ik}^{U_{jik}}\left(1-U_{jik}\right)Q_{ik}^{-U_{jik}}\left(-\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right) \\
&= \left\{U_{jik}P_{ik}^{U_{jik}-1}Q_{ik}^{1-U_{jik}} - P_{ik}^{U_{jik}}\left(1-U_{jik}\right)Q_{ik}^{-U_{jik}}\right\}\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}} \\
&= P_{ik}^{U_{jik}}Q_{ik}^{1-U_{jik}}\left\{\frac{U_{jik}}{P_{ik}} - \frac{(1-U_{jik})}{Q_{ik}}\right\}\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}} \\
&= P_{ik}^{U_{jik}}Q_{ik}^{1-U_{jik}}\left\{\frac{U_{jik}Q_{ik}-(P_{ik}-U_{jik}P_{ik})}{P_{ik}Q_{ik}}\right\}\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}} \\
&= P_{ik}^{U_{jik}}Q_{ik}^{1-U_{jik}}\left\{\frac{U_{jik}-P_{ik}}{P_{ik}Q_{ik}}\right\}\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}. \qquad (3.10)
\end{aligned}$$

Substituindo (3.10) em (3.8), obtemos

$$\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}} = \int_{I\!R^K}\left\{\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}. \qquad (3.11)$$

A função escore (3.5) fica

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}} &= \sum_{j=1}^{s} r_j \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\int_{I\!R^K}\left\{\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&= \sum_{j=1}^{s} r_j \int_{I\!R^K}\left\{\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\right\}\frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}d\boldsymbol{\theta},
\end{aligned}$$

que após aplicar a notação

$$g_j^*(\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta}|\mathbf{U}_{j..},\boldsymbol{\beta},\boldsymbol{\eta}) = \frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})} \qquad (3.12)$$

denotando a f.d.p da distribuição a posteriori do vetor de habilidades do $j - \acute{e}simo$

indivíduo (ou padrão de resposta), resulta em

$$\frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}} = \sum_{j=1}^{s} r_j \int_{I\!R^K} \left\{ \frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}} \left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}} \right) \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{3.13}$$

3.2.1 Aplicação ao Modelo Logístico de 3 Parâmetros

Quando consideramos o modelo logístico de 3 parâmetros, devemos usar as seguintes expressões:

$$\begin{aligned}
\frac{\partial P_{ik}}{\partial a_{ik}} &= D(1-c_{ik})(\theta_k - b_{ik})P_{ik}^* Q_{ik}^*, && (3.14)\\
\frac{\partial P_{ik}}{\partial b_{ik}} &= -Da_{ik}(1-c_{ik})P_{ik}^* Q_{ik}^*, && (3.15)\\
\frac{\partial P_{ik}}{\partial c_{ik}} &= Q_{ik}^*. && (3.16)
\end{aligned}$$

Assim, substituindo-se estas expressões em (3.13), obtemos as equações de estimação específicas para os parâmetros a_{ik}, b_{ik} e c_{ik}, dadas respectivamente por

$$\begin{aligned}
a_{ik} &: D(1-c_{ik}) \sum_{j=1}^{s} r_j \int_{I\!R^K} \left\{ \frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}} (\theta_k - b_{ik}) P_{ik}^* Q_{ik}^* \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, && (3.17)\\
b_{ik} &: -Da_{ik}(1-c_{ik}) \sum_{j=1}^{s} r_j \int_{I\!R^K} \left\{ \frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}} P_{ik}^* Q_{ik}^* \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, && (3.18)\\
c_{ik} &: \sum_{j=1}^{s} r_j \int_{I\!R^K} \left\{ \frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}} Q_{ik}^* \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0. && (3.19)
\end{aligned}$$

Podemos observar que estas equações não possuem solução explícita, sendo necessária portanto a aplicação de métodos iterativos para a obtenção das estimativas de máxima verossimilhança dos parâmetros dos itens, podendo ser o método de Newton-Raphson ou Escore de Fisher. Para isso será necessária a obtenção da derivada segunda da log-verossimilhança, ou seja, da Matriz Hessiana (ver Apêndice A.1). Note que teremos envolvido neste método a inversão de matrizes de ordem $3n$ para o caso do ML3, o que exige um grande esforço computacional. Para contornar este problema, apresentaremos a seguir uma proposta de Bock & Aitkin (1981).

3.3 Implementação

Apresentamos aqui as derivadas segundas da log-verossimilhança com relação aos parâmetros dos itens. Detalhes sobre a obtenção destas expressões podem ser encontradas no Apêndice A.1.

Usando a notação

$$\begin{aligned}
\mathbf{h}_{ik} &= \left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right) = \begin{pmatrix} D(1-c_{ik})(\theta_k - b_{ik}) \\ -Da_{ik}(1-c_{ik}) \\ \frac{1}{P^*_{ik}} \end{pmatrix} P^*_{ik}Q^*_{ik}, \text{ tem-se que} \\
\mathbf{H}_{iik} &= \left(\frac{\partial^2 P_{ik}}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}'_{ik}}\right) \\
&= \begin{pmatrix} D^2\left(\theta_k - b_{ik}\right)^2\left(1-c_{ik}\right)\left[1-2P^*_{ik}\right] & . & . \\ -D\left(1-c_{ik}\right)\left\{1+Da_{ik}\left(\theta_k - b_{ik}\right)\left[1-2P^*_{ik}\right]\right\} & D^2a^2_{ik}\left(1-c_{ik}\right)\left[1-2P^*_{ik}\right] & . \\ -D\left(\theta_k - b_{ik}\right) & Da_{ik} & 0 \end{pmatrix} P^*_{ik}Q^*_{ik}, \\
\mathbf{H}_{ii(j)} &= \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}'_{ik}} = \int_{I\!R^K}\left\{V_{jik}\mathbf{H}_{iik}\right\}g^*_j(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}, i=l, \text{ e} \\
\mathbf{H}_{il(j)} &= \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}'_{lk}} = \int_{I\!R^K}\left\{\left[V_{jlk}\mathbf{h}_{lk}\right]\left[V_{jik}\mathbf{h}'_{ik}\right]\right\}g^*_j(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}, i\neq l,
\end{aligned}$$

onde $V_{jik} = \frac{\left(U_{jik}-P_{ik}\right)}{P_{ik}Q_{ik}}$.

De forma análoga,

$$\mathbf{h}_{i(j)} = \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})/\partial \boldsymbol{\beta}_{ik}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})} = \int_{I\!R^K} V_{jik}\mathbf{h}_{ik}g^*_j(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}.$$

Então, fazendo

$$
\begin{aligned}
\mathbf{f}_1(\boldsymbol{\beta}_{ik}) &= \frac{\partial \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \beta_{ik}} \\
&= \sum_{j=1}^{s} r_j \int_{I\!R^K} V_{jik} \mathbf{h}_{ik} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}
$$

$$
\mathbf{H}(\boldsymbol{\beta}_{ik}, \boldsymbol{\beta}_{lk}) = \frac{\partial^2 \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik} \partial \boldsymbol{\beta}_{lk}^{'}} = \sum_{j=1}^{s} r_j \left[\mathbf{H}_{il(j)} - \mathbf{h}_{i(j)} \mathbf{h}_{l(j)}^{'} \right]
$$

e considerando que $\widehat{\boldsymbol{\beta}}^{(t)}$ é uma estimativa de $\boldsymbol{\beta}$ na iteração t, na iteração $t+1$ teremos

$$
\widehat{\boldsymbol{\beta}}^{(t+1)} = \widehat{\boldsymbol{\beta}}^{(t)} - \left[\mathbf{H}_{\mathbf{I}} \left(\widehat{\boldsymbol{\beta}}^{(t)} \right) \right]^{-1} \mathbf{f}_{\mathbf{I}} \left(\widehat{\boldsymbol{\beta}}^{(t)} \right),
$$

onde

$$
\mathbf{f}_{\mathbf{I}}(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{f}_1(\boldsymbol{\beta}_{11}) \\ \ldots \\ \mathbf{f}_1(\boldsymbol{\beta}_{n_1 1}) \\ \ldots \\ \mathbf{f}_1(\boldsymbol{\beta}_{1k}) \\ \ldots \\ \mathbf{f}_1(\boldsymbol{\beta}_{n_k k}) \end{pmatrix} \text{ e}
$$

$$
\mathbf{H}_{\mathbf{I}}(\boldsymbol{\beta}) = \begin{pmatrix}
\mathbf{H}(\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{11}) & \ldots & \mathbf{H}(\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{n_1 1}) & \mathbf{0} & \ldots & \ldots & \mathbf{0} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\mathbf{H}(\boldsymbol{\beta}_{n_1 1}, \boldsymbol{\beta}_{11}) & \ldots & \mathbf{H}(\boldsymbol{\beta}_{n_1 1}, \boldsymbol{\beta}_{n_1 1}) & \mathbf{0} & \ldots & \ldots & \mathbf{0} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\mathbf{0} & \ldots & \ldots & \mathbf{0} & \mathbf{H}(\boldsymbol{\beta}_{1k}, \boldsymbol{\beta}_{1k}) & \ldots & \mathbf{H}(\boldsymbol{\beta}_{1k}, \boldsymbol{\beta}_{n_k k}) \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\mathbf{0} & \ldots & \ldots & \mathbf{0} & \mathbf{H}(\boldsymbol{\beta}_{n_k k}, \boldsymbol{\beta}_{1k}) & \ldots & \mathbf{H}(\boldsymbol{\beta}_{n_k k}, \boldsymbol{\beta}_{n_k k})
\end{pmatrix}.
$$

Suposição Adicional

Para facilitar o processo de estimação podemos utilizar algumas suposições. Bock & Aitkin (1981) propuseram o que se denomina de *independência entre os itens*, de forma que

$$\mathbf{H}\left(\boldsymbol{\beta}_{ik},\boldsymbol{\beta}_{lk}\right)=\mathbf{0},\ \text{para } i\neq l,$$

o que torna a estrutura de $\mathbf{H_I}\left(\boldsymbol{\beta}\right)$ bloco diagonal, ou seja,

$$\mathbf{H_I}\left(\boldsymbol{\beta}\right)=diag\left\{\mathbf{H}\left(\boldsymbol{\beta}_{11},\boldsymbol{\beta}_{11}\right),...,\mathbf{H}\left(\boldsymbol{\beta}_{n_11},\boldsymbol{\beta}_{n_11}\right),...,\mathbf{H}\left(\boldsymbol{\beta}_{1k},\boldsymbol{\beta}_{1k}\right),...,\mathbf{H}\left(\boldsymbol{\beta}_{n_kk},\boldsymbol{\beta}_{n_kk}\right)\right\}.$$

Isto ameniza a exigência computacional, demandando um tempo bem menor no processo iterativo, pois agora a estimação dos parâmetros dos itens pode ser realizada individualmente, sendo necessária apenas a inversão de matrizes de ordem 3, ao invés de matrizes de ordem $3n$.

É importante observarmos que esta suposição é mais uma aproximação e não é equivalente à independência local, a qual está relacionada aos indivíduos, declarando que, uma vez fixada a habilidade, *as respostas* aos itens é que são independentes.

Sobre a métrica

É de extrema importância sabermos em que *métrica* os parâmetros dos itens foram estimados, pois é a partir do conhecimento da métrica que se inicia as interpretações. Neste bloco, os parâmetros populacionais são conhecidos, logo a métrica fica estabelecida. É importante observarmos que a métrica é inerente a cada população. Neste modelo as populações são caracterizadas pelas várias habilidades. Assim, os parâmetros dos itens estarão na métrica em que cada habilidade estiver, ou seja, serão estimados nas escalas (μ_k,σ_k^2), $k=1,...,K$.

Capítulo 4

Modelo Multivariado para uma Única População: Estimação dos Parâmetros Populacionais

Neste capítulo consideramos a estimação dos parâmetros populacionais supondo conhecidos os parâmetros dos itens. Na estimação conjunta este caso corresponde ao segundo bloco de estimação e na prática isto pode ocorrer quando os parâmetros dos itens já foram calibrados (estimados). Inicialmente tratamos do caso geral, sem especificar a distribuição e em seguida veremos o caso em que a distribuição do vetor de habilidades tem distribuição normal multivariada. Várias estruturas de covariâncias podem ser propostas para a relação entre as habilidades. Neste trabalho consideraremos a matriz de covariância não-estruturada, por ter uma estrutura generalizada, e apenas por simplicidade veremos o caso particular em que se tem uma estrutura uniforme.

4.1 Equação de Estimação: Caso Geral

As equações de estimação para os parâmetros populacionais são dadas por

$$\frac{\partial \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbf{0},$$

com $L(\boldsymbol{\beta}, \boldsymbol{\eta})$ dada por (3.4). Assim temos

$$\frac{\partial \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \sum_{j=1}^{s} r_j \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}, \tag{4.1}$$

onde

$$\begin{aligned}
\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} &= \frac{\partial}{\partial\boldsymbol{\eta}}\left\{\int_{I\!R^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}\right\} \qquad (4.2)\\
&= \int_{I\!R^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})\left(\frac{\partial}{\partial\boldsymbol{\eta}}g(\boldsymbol{\theta}|\boldsymbol{\eta})\right)d\boldsymbol{\theta}\\
&= \int_{I\!R^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})\left(\frac{\partial}{\partial\boldsymbol{\eta}}\log g(\boldsymbol{\theta}|\boldsymbol{\eta})\right)d\boldsymbol{\theta}. \qquad (4.3)
\end{aligned}$$

A ordem da derivada e da integral em (4.2) pôde ser permutada com base no Teorema da Convergência Dominada de Lebesgue [Chow & Teicher, 1978]. Assim, segue de (4.1)

$$\begin{aligned}
\frac{\partial\log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} &= \sum_{j=1}^{s} r_j \int_{I\!R^K}\left(\frac{\partial}{\partial\boldsymbol{\eta}}\log g(\boldsymbol{\theta}|\boldsymbol{\eta})\right)\frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}d\boldsymbol{\theta}\\
&= \sum_{j=1}^{s} r_j \int_{I\!R^K}\left(\frac{\partial}{\partial\boldsymbol{\eta}}\log g(\boldsymbol{\theta}|\boldsymbol{\eta})\right)g_j^*\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta} = \mathbf{0}. \qquad (4.4)
\end{aligned}$$

Algumas distribuições podem ser consideradas para as habilidades. Consideraremos a distribuição Normal por ser uma distribuição de maior interesse.

4.1.1 Habilidades Normalmente Distribuídas

No caso em que $\boldsymbol{\theta}$ tem distribuição Normal $K - variada$ com vetor de médias $\boldsymbol{\mu} = (\mu_1, ..., \mu_K)'$ e matriz de covariâncias Σ, teremos

$$\log g(\boldsymbol{\theta}|\boldsymbol{\eta}) = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)'\Sigma^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right),$$

que, após aplicar a derivada, fornece

$$\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\eta}}\left(\log|\Sigma|\right) - \frac{1}{2}\frac{\partial}{\partial\boldsymbol{\eta}}\left[\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)'\Sigma^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)\right].$$

Denotemos $\mathbf{V} = \Sigma^{-1} = (v_{kl})$, obtendo assim que

$$\begin{aligned}
\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_k} &= -\frac{1}{2}\frac{\partial}{\partial \mu_k}\left[(\boldsymbol{\theta}-\boldsymbol{\mu})'\,\mathbf{V}\,(\boldsymbol{\theta}-\boldsymbol{\mu})\right] \\
&= \mathbf{v}_k'\,(\boldsymbol{\theta}-\boldsymbol{\mu}), \quad k = 1, ..., K,
\end{aligned}$$

onde $\mathbf{v}_k$ é a $k-ésima$ coluna de $\mathbf{V}$. Usando notação matricial, podemos escrever as equações como

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu}} = \begin{pmatrix} \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_1} \\ \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_2} \\ \cdots \\ \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_k} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \cdots \\ \mathbf{v}_k' \end{pmatrix} (\boldsymbol{\theta}-\boldsymbol{\mu}) = \mathbf{V}\,(\boldsymbol{\theta}-\boldsymbol{\mu})\,. \tag{4.5}$$

A partir de (4.4) e (4.5) obtemos

$$\begin{aligned}
\mathbf{0} &= \sum_{j=1}^{s} r_j \int_{I\!R^K} \mathbf{V}\,(\boldsymbol{\theta}-\boldsymbol{\mu})\, g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \\
&= \sum_{j=1}^{s} r_j \mathbf{V} \left\{ \int_{I\!R^K} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} - \boldsymbol{\mu} \int_{I\!R^K} g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} \right\}.
\end{aligned}$$

Agora note que

$$\begin{aligned}
\int_{I\!R^K} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} &= \begin{pmatrix} \int_{I\!R^K} \theta_1 g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ \int_{I\!R^K} \theta_2 g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ \cdots \\ \int_{I\!R^K} \theta_K g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \int_{I\!R} \theta_1 \int_{I\!R^{K-1}} g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{(1)} d\theta_1 \\ \int_{I\!R} \theta_2 \int_{I\!R^{K-1}} g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{(2)} d\theta_2 \\ \cdots \\ \int_{I\!R} \theta_K \int_{I\!R^{K-1}} g_j^*(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{(K)} d\theta_K \end{pmatrix} = \\
&= \begin{pmatrix} \int_{I\!R} \theta_1 g_j^*(\theta_1)\, d\theta_1 \\ \int_{I\!R} \theta_2 g_j^*(\theta_2)\, d\theta_2 \\ \cdots \\ \int_{I\!R} \theta_K g_j^*(\theta_K)\, d\theta_K \end{pmatrix} = \begin{pmatrix} \widetilde{\mu}_{j1} \\ \widetilde{\mu}_{j2} \\ \cdots \\ \widetilde{\mu}_{jK} \end{pmatrix}
\end{aligned}$$

representa o vetor de médias a posteriori das habilidades. Desta forma, usando a notação

$$\widetilde{\boldsymbol{\mu}}_j = \int_{I\!R^K} \boldsymbol{\theta} g_j^* (\boldsymbol{\theta})\, d\boldsymbol{\theta} \qquad \text{e} \qquad \overline{\widetilde{\boldsymbol{\mu}}} = \frac{1}{N} \sum_{j=1}^{s} r_j \widetilde{\boldsymbol{\mu}}_j, \tag{4.6}$$

obteremos o seguinte resultado:

$$\begin{aligned} \mathbf{0} &= \sum_{j=1}^{s} r_j \mathbf{V} \left\{ \widetilde{\boldsymbol{\mu}}_j - \boldsymbol{\mu} \right\} = \mathbf{V} \left\{ \sum_{j=1}^{s} r_j \widetilde{\boldsymbol{\mu}}_j - N\boldsymbol{\mu} \right\} \\ &= N\mathbf{V} \left\{ \overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu} \right\}. \end{aligned} \tag{4.7}$$

São necessárias ainda as equações relativas aos parâmetros da estrutura de covariância, as quais serão apresentadas usando a matriz não-estruturada, que pode ser facilmente estendida para estruturas específicas. Assim, temos

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \sigma_{kl}} = -\frac{1}{2} \frac{\partial}{\partial \sigma_{kl}} (\log |\Sigma|) - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \left[\frac{\partial}{\partial \sigma_{kl}} \mathbf{V} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}).$$

Considerando $\delta_{kl} = \frac{\partial}{\partial \sigma_{kl}} \log |\Sigma|$ e $\mathbf{V}_{kl}^{(1)} = \frac{\partial}{\partial \sigma_{kl}} \mathbf{V}$ e usando a decomposição $(\boldsymbol{\theta} - \boldsymbol{\mu}) = \left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right) + \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right)$, teremos para a equação de estimação de σ_{kl}

$$\begin{aligned} 0 &= -\frac{1}{2} \sum_{j=1}^{s} r_j \int_{I\!R^K} \left(\delta_{kl} + (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{V}_{kl}^{(1)} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right) g_j^* (\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= -\frac{1}{2} \left(N\delta_{kl} + \sum_{j=1}^{s} r_j \int_{I\!R^K} \left[\left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right) + \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right) \right]' \mathbf{V}_{kl}^{(1)} \left[\left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right) + \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right) \right] \right) g_j^* (\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= -\frac{1}{2} \left(N\delta_{kl} + \sum_{j=1}^{s} r_j \int_{I\!R^K} \left[\left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right)' \mathbf{V}_{kl}^{(1)} \left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right) + \left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right)' \mathbf{V}_{kl}^{(1)} \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right) \right.\right. \\ &\quad \left.\left. + \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right)' \mathbf{V}_{kl}^{(1)} \left(\boldsymbol{\theta} - \overline{\widetilde{\boldsymbol{\mu}}}\right) + \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right)' \mathbf{V}_{kl}^{(1)} \left(\overline{\widetilde{\boldsymbol{\mu}}} - \boldsymbol{\mu}\right) \right] \right) g_j^* (\boldsymbol{\theta})\, d\boldsymbol{\theta} \end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}\left(N\delta_{kl}+\sum_{j=1}^{s} r_j\left[\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}\right.\right.\\
&\quad+\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)+\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}\\
&\quad\left.\left.+\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)\int_{I\!R^K}g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}\right]\right)\\
&= -\frac{1}{2}\left(N\delta_{kl}+\sum_{j=1}^{s} r_j\left[\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}+\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)\right]\right)\\
&= -\frac{1}{2}\left(N\delta_{kl}+\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)\sum_{j=1}^{s} r_j g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}+N\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)\right)\\
&= -\frac{N}{2}\left(\delta_{kl}+\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)\overline{g}_{\cdot}^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}+\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)\right),
\end{aligned}
$$

pois $\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)g_j^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}=\mathbf{0}$. Desta forma, temos

$$
\begin{aligned}
0 &= -\frac{N}{2}\left(\delta_{kl}+\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)\overline{g}_{\cdot}^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}+\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)\right)\\
&= \delta_{kl}+\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)\overline{g}_{\cdot}^{*}(\boldsymbol{\theta})\,d\boldsymbol{\theta}+\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\overline{\overline{\boldsymbol{\mu}}}-\boldsymbol{\mu}\right), \qquad (4.8)
\end{aligned}
$$

onde $\overline{g}_{\cdot}^{*}(\boldsymbol{\theta})=\sum_{j=1}^{s} r_j g_j^{*}(\boldsymbol{\theta})/N$. Como $\boldsymbol{\mu}$ é desconhecido, substituímos $\boldsymbol{\mu}$ por $\overline{\overline{\boldsymbol{\mu}}}$ usando a propriedade de invariância. E, finalmente usando (4.7) e (4.8), os EMV de $\boldsymbol{\mu}$ e σ_{kl} são obtidos resolvendo-se as equações

$$
\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \overline{\overline{\boldsymbol{\mu}}} \qquad (4.9)\\
\int_{I\!R^K}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)^{'}\mathbf{V}_{kl}^{(1)}\left(\boldsymbol{\theta}-\overline{\overline{\boldsymbol{\mu}}}\right)\overline{g}_{\cdot}^{*}(\boldsymbol{\theta})d\boldsymbol{\theta} &= -\delta_{kl}, \qquad (4.10)
\end{aligned}
$$

onde

$$
\begin{aligned}
\overline{\widetilde{\boldsymbol{\mu}}} &= \frac{1}{N}\sum_{j=1}^{s} r_j \widetilde{\boldsymbol{\mu}}_j, \text{ com } \widetilde{\boldsymbol{\mu}}_j = \int_{I\!R^K} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
\mathbf{V}_{kl}^{(1)} &= \frac{\partial \mathbf{V}}{\partial \sigma_{kl}}, \text{ com } \mathbf{V} = \Sigma^{-1}, \\
\delta_{kl} &= \frac{\partial \log |\Sigma|}{\partial \sigma_{kl}} \text{ e} \\
\overline{g}_{\cdot}^*(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{j=1}^{s} r_j g_j^*(\boldsymbol{\theta}), \text{ com } g_j^*(\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta}|\mathbf{U}_{j..}, \boldsymbol{\beta}, \boldsymbol{\eta}) = \frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\beta}) g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}.
\end{aligned}
$$

Portanto, para chegarmos às equações de verossimilhança para $\boldsymbol{\eta}$ basta que obtenhamos as expressões das derivadas primeiras de $\log|\Sigma|$ e de $\mathbf{V} = \Sigma^{-1}$.

A seguir, iremos considerar as derivadas para a matriz de covariância não-estruturada. Em particular, consideraremos o estudo da matriz de covariância Uniforme, a qual supõe iguais correlações entre as habilidades.

4.2 Estruturas de Covariância:

4.2.1 Matriz Não-Estruturada

De acordo com resultados obtidos em Mardia et al. (1979), temos que

$$
\mathbf{V}_{kl}^{(1)} = \frac{\partial \mathbf{V}}{\partial \sigma_{kl}} = \frac{\partial \Sigma^{-1}}{\partial \sigma_{kl}} = -\Sigma^{-1}\Delta_{kl}\Sigma^{-1}, \tag{4.11}
$$

onde Δ_{kl} é uma matriz $K \times K$ cujo elemento kl e lk é igual a 1 e os demais são iguais a zero. Além disso,

$$
\frac{\partial \log|\Sigma|}{\partial \Sigma} = \left(\frac{\partial \log|\Sigma|}{\partial \sigma_{kl}}\right) = (\delta_{kl}) = 2\Sigma^{-1} - \mathbf{D}_{\Sigma^{-1}},
$$

onde $\mathbf{D}_{\Sigma^{-1}}$ é uma matriz diagonal cujo $k-ésimo$ elemento é igual ao $k-ésimo$ elemento diagonal de Σ^{-1}. Portanto, cada elemento δ_{kl} é dado por

$$\delta_{kl} = \frac{\partial \log |\Sigma|}{\partial \sigma_{kl}} = \begin{cases} \sigma^{kk}; & k = l \\ 2\sigma^{kl}; & k \neq l, \end{cases} \tag{4.12}$$

onde σ^{kl} representa o $kl-ésimo$ elemento de Σ^{-1}.

Assim, as equações finais para a estimação dos parâmetros σ_{kl}, $1 \leq k \leq l \leq K$ pertencentes à matriz de covariância não-estruturada são obtidas ao substituir (4.11) e (4.12) em (4.10).

4.2.2 Matriz Uniforme

Essa estrutura, bastante explorada em estudos longitudinais, considera variâncias iguais para as K habilidades e correlação igual entre qualquer par (θ_k, θ_l). Assim, sua estrutura é dada por

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & ... & ... & \rho \\ \rho & 1 & \rho & ... & \rho \\ ... & ... & ... & ... & ... \\ \rho & ... & \rho & 1 & \rho \\ \rho & ... & ... & \rho & 1 \end{pmatrix},$$

podendo ser escrita como $\Sigma = \sigma^2 (1-\rho) \mathbf{I}_K + \sigma^2 \rho \mathbf{J}_K$, onde $\mathbf{J}_K = \mathbf{1}_K \mathbf{1}_K^{'}$ e $\mathbf{1}_K$ é o vetor $K \times 1$ de elementos unitários. Desta forma, pode-se verificar que

$$\begin{aligned} |\Sigma| &= \sigma^{2K} (1-\rho)^{K-1} (1 + (K-1)\rho), \\ \Sigma^{-1} &= \frac{1}{\sigma^2 (1-\rho)} \left(\mathbf{I}_K - \frac{\rho}{1 + (K-1)\rho} \mathbf{J}_K \right). \end{aligned}$$

Por simplicidade, usaremos a seguinte notação:

$$\alpha = \alpha(\rho) = (1-\rho), \quad \beta = \beta(\rho) = 1 + (K-1)\rho, \quad \delta = \delta(\rho) = (K-1)\alpha - \beta.$$

Assim, teremos

$$\begin{aligned} |\Sigma| &= \sigma^{2K}\alpha^{K-1}\beta, & (4.13)\\ \Sigma^{-1} &= \frac{1}{\sigma^2\alpha}\left(\mathbf{I}_K - \frac{\rho}{\beta}\mathbf{J}_K\right). & (4.14)\end{aligned}$$

Usaremos também a seguinte notação:

$$\mathbf{V}_\rho^{(1)} = \frac{\partial \mathbf{V}}{\partial \rho} = \left(\frac{\partial v_{ij}}{\partial \rho}\right) = \left(v_{ij\rho}^{(1)}\right)_{i,j\leq K}.$$

Agora de (4.13) e (4.14) segue

$$\begin{aligned} \frac{\partial \log|\Sigma|}{\partial \rho} &= -(K-1)\left[\frac{1}{\alpha} - \frac{1}{\beta}\right], & (4.15)\\ \mathbf{V}_\rho^{(1)} &= \frac{1}{\sigma^2}\left\{\frac{1}{\alpha^2}\mathbf{I}_K - \left(\frac{1}{\alpha\beta} - \frac{\rho\delta}{(\alpha\beta)^2}\right)\mathbf{J}_K\right\}. & (4.16)\end{aligned}$$

Podemos notar que esta estrutura apresenta uma forma geral dada por $\Sigma = \sigma^2\mathbf{R}$, onde $\mathbf{R}$ é uma matriz de correlação. Assim, temos que

$$\begin{aligned} \frac{\partial \log|\Sigma|}{\partial \sigma^2} &= \frac{\partial\left[\log \sigma^{2K}\right]}{\partial \sigma^2} = \frac{K}{\sigma^2}, & (4.17)\\ \frac{\partial \mathbf{V}}{\partial \sigma^2} &= \sigma^{-2}\mathbf{V}. & (4.18)\end{aligned}$$

Desta forma, as equações finais para a estimação dos parâmetros ρ e σ^2 são obtidas ao substituir (4.15), (4.16) e (4.17), (4.18) respectivamente em (4.10).

Note que no processo iterativo para a estimação de máxima verossimilhança dos parâmetros dos itens foram necessárias algumas suposições adicionais para que o esforço computacional fosse amenizado. Tais suposições não são necessárias no processo de estimação dos parâmetros populacionais, pois o número de parâmetros é relativamente pequeno.

Sobre a métrica

De forma análoga à estimação dos parâmetros dos itens, é de extrema importância sabermos em que *métrica* os parâmetros populacionais foram estimados. E como estamos tratando do caso em que os parâmetros dos itens são conhecidos, isto é, eles já foram estimados, o conhecimento dos parâmetros dos itens implica o conhecimento da métrica em que eles foram obtidos e, por conseqüência, os parâmetros populacionais serão estimados de acordo com a métrica dos parâmetros dos itens.

Assim, completamos o segundo bloco para o processo de estimação conjunta, no que se refere à obtenção das equações de estimação para os parâmetros populacionais. Como as equações de verossimilhança obtidas para as estruturas de covariância não apresentam soluções explícitas, precisaremos resolvê-las com a utilização de algum processo iterativo o qual será descrito no Apêndice A.2.

Capítulo 5

Estimação Conjunta: Parâmetros dos Itens e Populacionais

Aqui tratamos do caso mais comum, em que precisamos estimar os parâmetros dos itens e populacionais. Basicamente, o procedimento seria considerar como o conjunto das equações de estimação o conjunto obtido pela união das equações finais dos dois blocos apresentados anteriormente. Entretanto, na estimação conjunta surge um problema de métrica. Quando os parâmetros populacionais são conhecidos, os parâmetros dos itens são estimados na métrica dos parâmetros populacionais. Na situação oposta, os parâmetros populacionais são estimados na métrica dos parâmetros dos itens. Na estimação conjunta não há métrica estabelecida, de forma que o modelo torna-se *indeterminado*. No caso univariado, a forma mais usual de eliminar esse problema é fixar a escala de habilidade, em geral, $\theta \sim N(0,1)$. Então, de forma análoga, iremos fixar a escala de habilidade em cada um dos testes (ou disciplinas) em estudo.

Uma vez fixados os parâmetros populacionais em $\boldsymbol{\theta} \sim N_K(\mathbf{0}, \Sigma)$, com $\sigma_{11} = \sigma_{22} = ... = \sigma_{KK} = 1$, as expressões do segundo bloco devem ser adaptadas de modo a não incluir a estimação das médias e das variâncias das habilidades, restando apenas a estimação das covariâncias entre as habilidades. É interessante notar que, nesta situação, a matriz de correlação entre as habilidades é a própria matriz de covariância.

Vale a pena ressaltar que a medida de correlação não depende da métrica adotada para os diferentes testes. Por exemplo, se θ_k possui média μ_k e variância σ_k^2 e θ_l possui média μ_l e variância σ_l^2, então

$$Corr(\theta_k, \theta_l) = \frac{Cov(\theta_k, \theta_l)}{\sigma_k \sigma_l},$$

e se desejássemos que estas variáveis estivessem em outras métricas, poderíamos fazer as seguintes transformações lineares $\theta_k^* = \alpha_k + \beta_k \theta_k$ e $\theta_l^* = \alpha_l + \beta_l \theta_l$, de forma que

$$Corr(\theta_k^*, \theta_l^*) = \frac{Cov(\theta_k^*, \theta_l^*)}{\beta_k \beta_l \sigma_k \sigma_l} = \frac{\beta_k \beta_l Cov(\theta_k, \theta_l)}{\beta_k \beta_l \sigma_k \sigma_l} = Corr(\theta_k, \theta_l).$$

Na verdade, a medida de correlação é invariante a transformações lineares e portanto podemos sempre obter a métrica $(0, 1)$ pela transformação linear $\alpha_k = -\mu_k/\sigma_k$ e $\beta_k = 1/\sigma_k$ sem modificar a relação de dependência entre as habilidades.

5.1 Usando Suposições Adicionais

Na estimação conjunta, temos um número maior de parâmetros a estimar. Desta forma, faz-se necessário o uso de artifícios para facilitar o processo de estimação, como por exemplo algumas suposições. A primeira delas, já citada anteriormente, é a de independência entre os itens. Uma outra suposição é a de independência entre os itens e a população, cujo objetivo é tornar nula a derivada segunda da log-verossimilhança com relação a $\boldsymbol{\beta}_{ik}$ e $\boldsymbol{\eta}$.

Assumindo estas suposições, a estrutura da matriz Hessiana relativa à estimação conjunta dos parâmetros torna-se bloco diagonal, onde teremos matrizes de ordem 3 nos n primeiros blocos e escalares nos demais N blocos.

5.2 Aplicação do Algoritmo EM

O algoritmo EM proposto por Dempster, Laird & Rubin (1977) é um processo iterativo utilizado para obter estimativas de máxima verossimilhança na presença de observações (variáveis aleatórias) que podem ser vistas como dados incompletos (na presença de variáveis aleatórias não observadas). Como o próprio nome sugere, este processo consiste de dois passos em cada iteração: *Esperança* (E) e *Maximização* (M). No caso da TRI, podemos pensar no vetor de respostas $\mathbf{U}_{\ldots}$ como o vetor de dados incompletos (observado) e $(\mathbf{U}_{\ldots}, \boldsymbol{\theta})$ o vetor de dados completos (não-observado), e o objetivo é obter as estimativas de máxima verossimilhança dos parâmetros dos itens $\widehat{\boldsymbol{\beta}}$.

Inicialmente, Dempster et al. definiram o algoritmo EM para o caso em que a distribuição dos dados completos pertence à família exponencial, relaxando posteriormente esta hipótese em duas situações, onde eles consideram a família exponencial curvada (com restrições no espaço paramétrico) e o caso em que não há nenhuma relação com a família exponencial. Na TRI, somente o modelo logístico de 1 parâmetro (ou modelo Rasch) é membro da família exponencial regular, e isto simplifica o algoritmo, pois neste caso existe estatística suficiente simples para $\boldsymbol{\theta}$ sendo necessário então calcular a esperança da estatística suficiente $\boldsymbol{\theta}$ condicionada aos dados observados e aos parâmetros fixados (Passo E), e logo após substituir estas esperanças condicionais nas equações para a maximização (Passo M), realizando a estimação de máxima verossimilhança pelos métodos usuais (Newton-Raphson ou Escore de Fisher). Para o caso dos modelos de 2 e 3 parâmetros a distribuição conjunta de $f\left(\mathbf{U}_{\ldots}, \boldsymbol{\theta} | \boldsymbol{\beta}\right)$ é desconhecida e não pertence à família exponencial, sendo necessária, portanto, uma adaptação para a aplicação da terceira definição do algoritmo EM. Descreveremos agora a adaptação do nosso modelo para a aplicação do algoritmo.

Usando (3.13) e considerando

$$f\left(\boldsymbol{\theta}\right) \equiv f_{ik}\left(\boldsymbol{\theta}\right) = \sum_{j=1}^{s} r_j g_j^*(\boldsymbol{\theta}) \quad \text{e} \quad r_{ik}\left(\boldsymbol{\theta}\right) = \sum_{j=1}^{s} r_j U_{jik} g_j^*(\boldsymbol{\theta}),$$

que representam, respectivamente, o número esperado de indivíduos respondendo ao item i no teste K e o número esperado de indivíduos respondendo corretamente ao item i no teste K, podemos obter

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}} &= \int_{I\!R^K}\left\{\frac{1}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\left[\sum_{j=1}^{s} r_j U_{jik} g_j^*(\boldsymbol{\theta}) - P_{ik}\sum_{j=1}^{s} r_j g_j^*(\boldsymbol{\theta})\right]\right\} d\boldsymbol{\theta} \\
&= \int_{I\!R^K}\left\{\frac{1}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\left[r_{ik}\left(\boldsymbol{\theta}\right) - P_{ik} f\left(\boldsymbol{\theta}\right)\right]\right\} d\boldsymbol{\theta}. \qquad (5.1)
\end{aligned}$$

Desta maneira, $\mathbf{f} = (f_{ik})$ e $\mathbf{r} = (r_{ik})$ correspondem às quantidades não-observadas do algoritmo EM.

Podemos notar que a integral em (5.1) não apresenta solução analítica. Na literatura existem vários métodos disponíveis para sua solução. O que tem sido bastante utilizado na TRI é o método de *Hermite-Gauss* ou *método de quadratura gaussiana* [Stroud & Secrest, 1966]. Este método considera uma aproximação baseada num conjunto de Q pontos $\overline{\theta}_l$ com os respectivos pesos A_l, $l = 1, ..., Q$.

Portanto, a equação na forma de quadratura é dada por

$$\frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}} = \sum_{1\le l_p \le Q}\left\{\frac{1}{P_i\left(\theta_{l_k}\right)Q_i\left(\theta_{l_k}\right)}\left(\frac{\partial P_i\left(\theta_{l_k}\right)}{\partial \boldsymbol{\beta}_{ik}}\right)\left[r_{ik}\left(\overline{\theta}_{l_p}\right) - P_i\left(\theta_{l_k}\right) f\left(\overline{\theta}_{l_p}\right)\right]\right\} A_l,$$

onde $p = 1, ..., k, ..., K$ e $Q = \prod_{k=1}^{K} q_k$, em que q_k representa o número de pontos de quadratura na $k - \acute{e}sima$ dimensão.

Desta maneira, os passos E e M podem ser descritos como a seguir.

Passo E

Usar os pontos de quadratura $\overline{\theta}_l = (\theta_{l_1}, \theta_{l_2}, ..., \theta_{l_K})$, os pesos A_l, $l = 1, ..., Q$, estimativas iniciais dos parâmetros dos itens, $\widehat{\boldsymbol{\beta}}_i$, $i = 1, ..., n$, e dos parâmetros populacionais, $\widehat{\boldsymbol{\eta}}$, para gerar $g_j^*(\overline{\boldsymbol{\theta}}_l)$ e, posteriormente, $f_{lik} = f_{ik}\left(\overline{\boldsymbol{\theta}}_l\right)$ e $r_{lik} = r_{ik}\left(\overline{\boldsymbol{\theta}}_l\right)$, $i = 1, ..., n$.

Passo M

1. Usar os pontos de quadratura e $g_j^*(\overline{\boldsymbol{\theta}}_l)$ para obter $\widehat{\boldsymbol{\eta}}$ por (4.9) e (4.10).
2. Com $\mathbf{r}$, $\mathbf{f}$ e $\widehat{\boldsymbol{\eta}}$ obtidos no passo E e M-1, resolver as equações de estimação para $\boldsymbol{\beta}_{ik}$, $i \in I_k$, $k = 1, ..., K$, usando o algoritmo de Newton-Raphson ou Escore de Fisher (Ver Apêndice A.3).

Estes passos compõem cada iteração do algoritmo EM, os quais devem ser repetidos até que algum critério de parada seja alcançado. Vale notar que este algoritmo pode ser utilizado na situação em que apenas os parâmetros dos itens são de interesse e neste caso o passo M-1 torna-se desnecessário.

Estimativas iniciais

No processo de estimação, uma atenção especial deve ser dada para a utilização de estimativas iniciais para os parâmetros. Alguns autores alertam que más estimativas iniciais podem revelar irregularidades na sequência de estimativas produzidas pelo Método de Newton-Raphson [Cordeiro, 1999].

Algumas propostas são apresentadas em Andrade, Tavares Valle (2000) para o caso de modelos univariados, e como o nosso modelo não apresenta itens comuns, parece ser razoável que utilizemos tais propostas. Richardson (1936) e Tucker (1946) mostraram que ao utilizar a Função de Resposta ao Item-FRI Normal, os seguintes valores iniciais podem ser utilizados:

$$\widehat{a}_{ik}^{(0)} = \frac{\rho_{T_k,U_{ik}}}{\sqrt{1-\rho^2_{T_k,U_{ik}}}}, \quad -1 < \rho_{T_k,U_{ik}} < 1,$$

onde $\rho_{T_k,U_{ik}}$ é o coeficiente ponto bisserial, o qual é estimado pelo coeficiente de correlação de Pearson entre os escores do teste k, denotado por T_k, e as resposta ao item i no teste k, U_{ik}. Tucker (1946) expressou o parâmetro de dificuldade do item i no teste k da teoria clássica de itens π_{ik} (proporção verdadeira de respostas corretas do item i no teste k)

como

$$\pi_{ik} = \Phi\left(-\upsilon_{ik}\right), \ \upsilon_{ik} = b_{ik}\ \rho_{T_k,U_{ik}},$$

onde Φ é a função de distribuição associada à distribuição normal-padrão. Um "chute" razoável para o parâmetro de acerto ao acaso é $c_{ik} = 1/m_{ik}$, onde m_{ik} é o número de alternativas possíveis do item i no teste k. Para os Parâmetros Populacionais, os escores padronizados em cada teste e as correlações entre estes escores também são estimativas razoáveis para o vetor de médias e matriz de correlação das habilidades em estudo.

Ressaltamos que, apesar de não estarmos utilizando a FRI normal nem o valor de $D = 1,7$ para tornar os modelos logístico próximos do modelo normal, possivelmente utilizaremos no processo iterativo tais propostas. Ressaltamos também que uma proposta para evitar estimativas irregulares seria a utilização do Método Iterativo de Quasi-Newton, o qual considera aproximações para as derivadas segundas apresentando menor dependência dos valores iniciais.

Para finalizar este capítulo, é importante fazermos algumas considerações. Primeiramente, alertamos que para alguns conjuntos de dados pode haver uma certa instabilidade no processo de estimação por máxima verossimilhança, levando à má estimação de alguns parâmetros de itens. Pode ocorrer, por exemplo, $a_{ik} < 0$ ou $c_{ik} \notin [0,1]$. Além disso, para os itens que são respondidos corretamente, ou incorrretamente, por todos os indivíduos, os parâmetros divergem para $+\infty$ ou $-\infty$, respectivamente. De forma análoga ocorre com as *estimativas das habilidades* de indivíduos que respondem todos os itens corrretamente, ou incorretamente. Para contornar estes problemas, métodos bayesianos são propostos, os quais atribuem distribuições a priori para os parâmetros. Um segundo ponto é que até o momento descrevemos apenas a estimação dos parâmetros dos itens e populacionais, não havendo interesse na estimação das habilidades individuais dos indivíduos. No entanto, há aplicações cujo interesse é estimar tais habilidades, como nos processos seletivos.

Diante das considerações acima, descreveremos brevemente no próximo capítulo os procedimentos bayesianos para a estimação dos parâmetros dos itens, populacionais, bem como da estimação bayesiana das habilidades individuais.

Capítulo 6

Estimação Bayesiana: Parâmetros dos Itens, Populacionais e Habilidades

Dentre as várias propostas para a estimação bayesiana, a mais utilizada é a *Estimação Bayesiana Marginal* [Mislevy, 1986], que é uma generalização da estimação por Máxima Verossimilhança Marginal em que distribuições a priori para os parâmetros de interesse são atribuídas e estimativas são obtidas através de alguma característica da função de distribuição a posteriori, tal como a moda e a média.

Neste procedimento, consideramos $(\boldsymbol{\beta}, \boldsymbol{\eta})$ variáveis aleatórias independentes e contínuas, com distribuições especificadas. Assim, os parâmetros $\boldsymbol{\beta}_{ik}$, $i \in I_k$, $k = 1, ..., K$, terão a função densidade denotada por $g\left(\boldsymbol{\beta}_{ik}|\boldsymbol{\tau}\right)$ e a função densidade de $\boldsymbol{\eta}$ por $h\left(\boldsymbol{\eta}|\boldsymbol{\varphi}\right)$, então a densidade conjunta desses parâmetros é dada por

$$f\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\varphi}\right) = f\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) g\left(\boldsymbol{\theta}|\boldsymbol{\eta}\right) h\left(\boldsymbol{\eta}|\boldsymbol{\varphi}\right).$$

Para fazer inferências com relação a $\boldsymbol{\beta}$ e $\boldsymbol{\eta}$, devemos nos basear na *distribuição a posteriori* de $\boldsymbol{\beta}$ e $\boldsymbol{\eta}$, que é dada por

$$\begin{aligned}
f^*\left(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{U}...\right) &= C \int_{I\!R^K} P(\mathbf{U}_{...}|\boldsymbol{\theta}, \boldsymbol{\beta}) f\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) g(\boldsymbol{\theta}|\boldsymbol{\eta}) h\left(\boldsymbol{\eta}|\boldsymbol{\varphi}\right) d\boldsymbol{\theta} \\
&= C f\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) h\left(\boldsymbol{\eta}|\boldsymbol{\varphi}\right) \left\{ \int_{I\!R^K} P(\mathbf{U}_{...}|\boldsymbol{\theta}, \boldsymbol{\beta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \right\} \\
&\propto L(\boldsymbol{\beta}, \boldsymbol{\eta}) f\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) h\left(\boldsymbol{\eta}|\boldsymbol{\varphi}\right),
\end{aligned}$$

onde C é uma constante e $L(\boldsymbol{\beta}, \boldsymbol{\eta}) \equiv P(\mathbf{U}_{...}; \boldsymbol{\beta}, \boldsymbol{\eta})$.

Para estimar $\boldsymbol{\beta}$ e $\boldsymbol{\eta}$ podemos escolher alguma característica de $f^*(\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{U}...)$, como por exemplo a média ou a moda, as quais são mais utilizadas. Neste trabalho consideraremos a moda a posteriori, ou seja, o valor de $\boldsymbol{\beta}$ e $\boldsymbol{\eta}$ que maximiza a posteriori marginal. Desta maneira, temos que

$$\log f^*(\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{U}...) = \text{Const} + \log L(\boldsymbol{\beta},\boldsymbol{\eta}) + \log f(\boldsymbol{\beta}|\boldsymbol{\tau}) + \log h(\boldsymbol{\eta}|\boldsymbol{\varphi}),$$

onde o primeiro termo é uma constante. Derivando esta expressão em relação a $\boldsymbol{\beta}_{ik}$, $i \in I_k, \;\; k = 1,...,K$ e $\boldsymbol{\eta}$, obtemos

$$\begin{aligned}
\frac{\partial \log f^*(\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{U}...)}{\partial \beta_{ik}} &= \frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}} + \frac{\partial \log f(\boldsymbol{\beta}|\boldsymbol{\tau})}{\partial \beta_{ik}} = \mathbf{0}, \\
\frac{\partial \log f^*(\boldsymbol{\beta},\boldsymbol{\eta}|\mathbf{U}...)}{\partial \boldsymbol{\eta}} &= \frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} + \frac{\partial \log h(\boldsymbol{\eta}|\boldsymbol{\varphi})}{\partial \boldsymbol{\eta}} = \mathbf{0}.
\end{aligned}$$

Note que, exceto pela segunda parcela, estas equações são análogas às obtidas em (3.13) e (4.4), respectivamente, pois a abordagem bayesiana apenas adiciona uma nova parcela relativa à distribuição a priori associada aos parâmetros de interesse. As segundas parcelas dependerão das prioris adotadas para os parâmetros, podendo ser baseadas em prioris pré-especificadas (Solução Bayesiana) ou estimadas através dos dados (Solução Bayesiana Empírica).

As distribuições dos parâmetros dos itens devem levar em consideração as limitações de seus valores. O parâmetro a_{ik} só pode assumir valores positivos, b_{ik} qualquer valor real e c_{ik} deve pertencer ao intervalo $[0,1]$. Com base nestas limitações, as propostas mais comuns são as distribuições Log-Normal ou qui-quadrado para a_{ik}, a distribuição Normal para b_{ik} e a Beta para c_{ik}. Alguns autores, como Baker (1992), preferem utilizar expressões para a estimação de $\alpha_{ik} = \log a_{ik}$ (para o caso de $a_{ik} \sim$Log-Normal), o qual tem distribuição Normal, sugerindo a utilização da propriedade de invariância do estimador de máxima verossimilhança para obter $\hat{a}_{ik}$ pela transformação $\hat{a}_{ik} = \exp(\hat{\alpha}_{ik})$. Swaminatham & Gifford (1986) propuseram uma reparemetrização dos parâmetros da distribuição de c_{ik} de forma que a média pudesse ser interpretada como a probabilidade de acerto por indivíduos de baixa habilidade.

Em relação aos parâmetros populacionais não existem problemas de instabilidade como na estimação dos parâmetros dos itens por máxima verossimilhança. No entanto, se algum conhecimento prévio estiver disponível, pode ser de interesse considerar uma distribuição a priori para os parâmetros populacionais. Pode-se, por exemplo, considerar que $\boldsymbol{\theta} \sim N_K(\boldsymbol{\mu}, \Sigma)$, onde $\boldsymbol{\mu} \sim N_K\left(\boldsymbol{\mu}_\mu, \Sigma_\mu\right)$ e as prioris para a matriz de covariância Σ dependem da estrutura adotada.

Não nos estendemos em maiores detalhes sobre os procedimentos bayesianos para a estimação dos parâmetros dos itens e populacionais por acreditar que existem boas referências sobre o assunto, como por exemplo Andrade, Tavares & Valle (2000) e Tavares (2001).

6.1 Estimação das Habilidades

De acordo com a Seção 2.3, vimos que uma vez conhecido os parâmetros dos itens, as habilidades individuais podem ser obtidas por Máxima Verossimilhança ou através de alguma característica da distribuição de θ_j condicionado ao vetor de respostas. De forma análoga, no modelo multivariado pode-se usar (3.2) para obter estimativas por máxima verossimilhança. No entanto, o procedimento por máxima verossimilhança ignora a informação oriunda das várias habilidades. Por conta disso, a estimação de $\boldsymbol{\theta}_j$ será feita com base na distribuição condicional de $\boldsymbol{\theta}_j$, dado o conjunto de respostas do $j-\acute{e}simo$ indivíduo, pois desta forma é possível utilizar a informação contida na estrutura de covariância entre as K habilidades.

6.1.1 Moda da Posteriori de $\boldsymbol{\theta}_j$

Para obter a moda da distribuição a posteriori de $\boldsymbol{\theta}_j$ (ver (3.12)) observamos que

$$\log g_j^*(\boldsymbol{\theta}_j) = \log \mathbf{P}(\boldsymbol{\theta}_j|\mathbf{U}_{j..}, \boldsymbol{\beta}, \boldsymbol{\eta}) \propto \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta}) + \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta}).$$

Portanto,

$$\frac{\partial \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \propto \frac{\partial \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \boldsymbol{\theta}_j} + \frac{\partial \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j} = \mathbf{0}. \tag{6.1}$$

Agora notemos que

$$\log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta}) = \sum_{k=1}^{K} \sum_{i \epsilon I_k} \log P(U_{jik}|\theta_{jk}, \boldsymbol{\beta}_{ik})$$

e

$$\frac{\partial \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \boldsymbol{\theta}_j} = \left(\frac{\partial \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \theta_{j1}}, ..., \frac{\partial \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \theta_{jK}} \right)'.$$

Acontece que, de forma análoga ao desenvolvimento da expressão entre chaves em (3.7), podemos obter

$$\frac{\partial \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \theta_{jk}} = \frac{(U_{jik} - P_{ik})}{P_{ik} Q_{ik}} \left(\frac{\partial P_{ik}}{\partial \theta_{jk}} \right) = V_{jik} \left(\frac{\partial P_{jik}}{\partial \theta_{jk}} \right),$$

onde P_{jik} é dada por (3.1).

Além disso, temos

$$\frac{\partial P_{jik}}{\partial \theta_{jk}} = -\frac{\partial P_{jik}}{\partial b_{ik}} = Da_{ik}(1 - c_{ik}) P_{ik}^* Q_{ik}^* = \alpha_{ik} P_{jik}^* Q_{jik}^*,$$

em que P^*_{jik} é a probabilidade associada ao ML2. Desta forma,

$$\frac{\partial \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \theta_{jk}} = \alpha_{ik} V_{jik} P^*_{jik} Q^*_{jik}. \tag{6.2}$$

Verificamos também que

$$\begin{aligned} \frac{\partial \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\theta}_{jk}} \left(\boldsymbol{\theta}_j - \boldsymbol{\mu}\right)' \Sigma^{-1} \left(\boldsymbol{\theta}_j - \boldsymbol{\mu}\right) \\ &= -\Sigma^{-1} \left(\boldsymbol{\theta}_j - \boldsymbol{\mu}\right). \end{aligned} \tag{6.3}$$

Assim, de (6.2) e (6.3) chegamos à equação de estimação para $\boldsymbol{\theta}_j$, $j = 1, ..., N$. Como (6.1) não possui solução explícita, será necessário aplicar algum processo iterativo.

Aplicação do Método de Newton-Raphson

Este procedimento requer a derivada segunda de $g^*_j(\boldsymbol{\theta}_j)$. A seguir apresentaremos as expressões principais, reservando o Apêndice A.4 para a apresentação dos detalhes.

Temos

$$\frac{\partial^2 \log g^*_j(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} \propto \frac{\partial^2 \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} + \frac{\partial^2 \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j}, \tag{6.4}$$

onde

$$\frac{\partial^2 \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \theta_{jk} \partial \theta_{jl}} = \begin{cases} \alpha^2_{ik} V_{jik} \left[-V_{jik} P^*_{jik} Q^*_{jik} + \left(1 - 2P^*_{jik}\right)\right] P^*_{jik} Q^*_{jik}, & k = l, \\ 0, & k \neq l, \end{cases}$$

e, de (6.3),

$$\frac{\partial^2 \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} = -\Sigma^{-1}.$$

Agora, usando a notação

$$\mathbf{f}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j\right) = \frac{\partial \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}, \quad \mathbf{H}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j\right) = \frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'},$$

e sendo $\hat{\boldsymbol{\theta}}_j^{(t+1)}$ a estimativa de $\boldsymbol{\theta}_j$ na iteração t, teremos na iteração $t+1$

$$\hat{\boldsymbol{\theta}}_j^{(t+1)} = \hat{\boldsymbol{\theta}}_j^{(t)} - \left[\mathbf{H}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_j^{(t)}\right)\right]^{-1} \mathbf{f}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_j^{(t)}\right).$$

6.1.2 Média da Posteriori de $\boldsymbol{\theta}_j$

Uma outra forma de estimação que tem a vantagem de ser calculada diretamente, sem recorrer à aplicação de métodos iterativos, é obtida ao se estimar $\boldsymbol{\theta}_j$ através da média da posteriori $g_j^*(\boldsymbol{\theta}_j)$. Desta forma, se lembrarmos de (3.12) e (3.3) verificamos que o estimador será obtido por

$$\begin{aligned} \hat{\boldsymbol{\theta}}_j &= \mathbb{E}\left(\boldsymbol{\theta}|\mathbf{U}_{j..}, \boldsymbol{\beta}, \boldsymbol{\eta}\right) \\ &= \int_{\mathbb{R}^K} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \\ &= \frac{\int_{\mathbb{R}^K} \boldsymbol{\theta} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\beta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\beta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta}}. \end{aligned} \tag{6.5}$$

Note que, além de não ser necessária a aplicação de métodos iterativos, esta quantidade já aparece no processo de estimação dos parâmetros populacionais, através da primeira expressão em (4.6). Portanto, as habilidades individuais já podem estar disponíveis ao final da primeira etapa do processo de estimação via algoritmo EM.

As aplicações a dados simulados e a dados reais serão considerados no próximo capítulo.

Capítulo 7

Simulação e Aplicação a Dados Reais

Apresentaremos neste capítulo a aplicação a dados simulados e a dados reais com o intuito de observar e validar os resultados obtidos pela teoria desenvolvida nos capítulos anteriores.

As simulações aqui apresentadas serão relativas ao caso mais realista, que considera a *Estimação Conjunta* (*Parâmetros Populacionais* e de *Itens).* Na Seção 7.1 consideraremos a estimação dos parâmetros nas situações em que a distribuição das habilidades possui matriz de covariância *Uniforme* e *Não-Estruturada.* Uma aplicação aos dados do Vestibular 2000 - UFC é apresentada na Seção 7.2, onde procuramos explorar as duas estruturas de covariância acima citadas.

Será considerada para o estudo de simulação a situação em que $N = 1000$ indivíduos são submetidos a $K = 5$ testes relacionados a 5 distintas áreas do conhecimento. De acordo com o que foi dito na Seção 3.1.1, iremos utilizar o ML3 para a probabilidade de uma determinada resposta a qualquer um dos itens e adotaremos o fator $D = 1,7$ (escala normal). Em relação à distribuição $N_K(\boldsymbol{\mu}, \Sigma)$, temos que o vetor de médias será $\boldsymbol{\mu} = (0, ..., 0)'$ e a matriz de covariância terá variâncias $\sigma_{11} = \sigma_{22} = ... = \sigma_{KK} = 1$, uma vez que na estimação conjunta o problema de indeterminação é resolvido ao se fixar os parâmetros de locação e escala em valores quaisquer (ver Cap. 5). Na Tabela 7.1 mostram-se os verdadeiros valores das covariâncias entre as habilidades. Devemos ressaltar que os valores gerados para as habilidades sofreram uma transformação do tipo $\boldsymbol{\theta}_j^* = \Sigma^{1/2}\mathbf{S}^{-1/2}(\boldsymbol{\theta}_j - \mathbf{m})$, $j = 1, ..., N$, onde $\mathbf{m}$ é o vetor de médias da amostra gerada, $\mathbf{S}$ é a matriz de covariância amostral e Σ é a matriz de covariância populacional. Esse procedimento gerou valores para as habilidades com os parâmetros desejados e demonstrou-se bastante satisfatório numericamente.

No que diz respeito aos $K = 5$ testes, consideramos a existência de 24 itens em cada um. Os verdadeiros valores para os parâmetros a, b e c são apresentados na Tabela 7.2, onde os 24 primeiros itens correspondem ao *Teste 1,* os 24 seguintes ao *Teste 2* e assim por diante. Deve-se observar que os parâmetros estão na escala normal e por isso os valores de a_{ik} variam de $0,6$ (baixa discriminação) a $1,4$ (alta discriminação), os parâmetros b_{ik} variam de $-2,0$ (item mais fácil) a $2,0$ (item mais difícil) e os valores $0,20$ e $0,25$ são considerados para o parâmetro c_{ik}. Pode-se observar ainda que itens de várias características compõem cada teste de forma bastante representativa.

Devemos esclarecer que nas simulações não utilizamos os "chutes iniciais" referidos no Capítulo 5 e sim valores obtidos a partir de pertubações aleatórias e uniformes em intervalos que variam de acordo com o tipo de parâmetro de item: $\varepsilon_a \in (-0,2; 0,2)$, $\varepsilon_b \in (-0,5; 0,5)$ e $\varepsilon_c \in (-0,1; 0,1)$. Somente assim foi possível previnir (ou evitar) irregularidades no processo iterativo.

Nas simulações foi utilizada a linguagem de programação matricial orientada a objetos Ox versão 2.20 [Doornik, 1999]. O programa desenvolvido para as simulações pode ser solicitado através do e-mail: gilberto_matos@uol.com.br.

Tabela 7.1: Valores dos parâmetros populacionais usados nas simulações.

Vetor de Médias	$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$
Estrutura de Covariância	
Uniforme:	$\sigma^2 = 1, \quad \rho = 0,25$ e $0,75$
Não-Estruturada:	
$\Sigma =$	$\begin{bmatrix} 1 & 0,45 & 0,75 & 0,75 & 0,75 \\ . & 1 & 0,75 & 0,35 & 0,65 \\ . & . & 1 & 0,75 & 0,75 \\ . & . & . & 1 & 0,75 \\ . & . & . & . & 1 \end{bmatrix}$

Tabela 7.2: Valores dos parâmetros dos itens.

T	Item i	a_i	b_i	c_i	T	Item i	a_i	b_i	c_i	T	Item i	a_i	b_i	c_i
1	1	1,0	-2	0,25	2	25	1,0	-2	0,25	3	49	1,0	-2	0,25
1	2	1,4	-2	0,20	2	26	1,4	-2	0,20	3	50	1,4	-2	0,20
1	3	1,4	-2	0,25	2	27	1,4	-2	0,25	3	51	1,4	-2	0,25
1	4	0,6	-1	0,20	2	28	0,6	-1	0,20	3	52	0,6	-1	0,20
1	5	0,6	-1	0,25	2	29	0,6	-1	0,25	3	53	0,6	-1	0,25
1	6	1,0	-1	0,20	2	30	1,0	-1	0,20	3	54	1,0	-1	0,20
1	7	1,0	-1	0,25	2	31	1,0	-1	0,25	3	55	1,0	-1	0,25
1	8	1,4	-1	0,20	2	32	1,4	-1	0,20	3	56	1,4	-1	0,20
1	9	1,4	-1	0,25	2	33	1,4	-1	0,25	3	57	1,4	-1	0,25
1	10	0,6	0	0,20	2	34	0,6	0	0,20	3	58	0,6	0	0,20
1	11	0,6	0	0,25	2	35	0,6	0	0,25	3	59	0,6	0	0,25
1	12	1,0	0	0,20	2	36	1,0	0	0,20	3	60	1,0	0	0,20
1	13	1,0	0	0,25	2	37	1,0	0	0,25	3	61	1,0	0	0,25
1	14	1,4	0	0,20	2	38	1,4	0	0,20	3	62	1,4	0	0,20
1	15	1,4	0	0,25	2	39	1,4	0	0,25	3	63	1,4	0	0,25
1	16	0,6	1	0,20	2	40	0,6	1	0,20	3	64	0,6	1	0,20
1	17	0,6	1	0,25	2	41	0,6	1	0,25	3	65	0,6	1	0,25
1	18	1,0	1	0,20	2	42	1,0	1	0,20	3	66	1,0	1	0,20
1	19	1,0	1	0,25	2	43	1,0	1	0,25	3	67	1,0	1	0,25
1	20	1,4	1	0,20	2	44	1,4	1	0,20	3	68	1,4	1	0,20
1	21	1,4	1	0,25	2	45	1,4	1	0,25	3	69	1,4	1	0,25
1	22	0,6	2	0,20	2	46	0,6	2	0,20	3	70	0,6	2	0,20
1	23	0,6	2	0,25	2	47	0,6	2	0,25	3	71	0,6	2	0,25
1	24	1,0	2	0,20	2	48	1,0	2	0,20	3	72	1,0	2	0,20

T: Teste

Tabela 7.2: Cont. .

T	Item i	a_i	b_i	c_i	T	Item i	a_i	b_i	c_i
4	73	1,0	-2	0,25	5	97	1,0	-2	0,25
4	74	1,4	-2	0,20	5	98	1,4	-2	0,20
4	75	1,4	-2	0,25	5	99	1,4	-2	0,25
4	76	0,6	-1	0,20	5	100	0,6	-1	0,20
4	77	0,6	-1	0,25	5	101	0,6	-1	0,25
4	78	1,0	-1	0,20	5	102	1,0	-1	0,20
4	79	1,0	-1	0,25	5	103	1,0	-1	0,25
4	80	1,4	-1	0,20	5	104	1,4	-1	0,20
4	81	1,4	-1	0,25	5	105	1,4	-1	0,25
4	82	0,6	0	0,20	5	106	0,6	0	0,20
4	83	0,6	0	0,25	5	107	0,6	0	0,25
4	84	1,0	0	0,20	5	108	1,0	0	0,20
4	85	1,0	0	0,25	5	109	1,0	0	0,25
4	86	1,4	0	0,20	5	110	1,4	0	0,20
4	87	1,4	0	0,25	5	111	1,4	0	0,25
4	88	0,6	1	0,20	5	112	0,6	1	0,20
4	89	0,6	1	0,25	5	113	0,6	1	0,25
4	90	1,0	1	0,20	5	114	1,0	1	0,20
4	91	1,0	1	0,25	5	115	1,0	1	0,25
4	92	1,4	1	0,20	5	116	1,4	1	0,20
4	93	1,4	1	0,25	5	117	1,4	1	0,25
4	94	0,6	2	0,20	5	118	0,6	2	0,20
4	95	0,6	2	0,25	5	119	0,6	2	0,25
4	96	1,0	2	0,20	5	120	1,0	2	0,20

T: Teste

7.1 Aplicação a Dados Simulados

Nesta seção veremos os resultados obtidos para a estimação conjunta dos parâmetros populacionais e de itens, situação esta que ocorre com maior freqüência. É importante lembrar que a situação onde se deseja estimar os parâmetros populacionais e/ou as habilidades individuais com os parâmetros dos itens conhecidos está tornando-se uma prática cada vez mais comum devido à criação de banco de itens, como, e.g, o Banco Nacional de Itens do INEP/MEC.

Para simular a situação em que a correlação entre qualquer par de habilidades é sempre a mesma, utilizamos uma versão do programa denominado LONGIRT desenvolvido por Tavares (2001).

Parâmetros Populacionais

Observando a Tabela 7.3, vemos que as estimativas médias obtidas a partir de $R = 500$ réplicas estão muito próximas dos verdadeiros valores, além de apresentarem pequena variabilidade. Quando nos restringimos à estrutura uniforme vemos que a estimativa foi menos precisa para $\rho = 0,25$. Para a matriz não-estruturada podemos notar que, apesar de algumas verdadeiras correlações serem iguais, as estimativas médias não apresentaram resultados muito similares. Por exemplo, para correlações iguais a $0,75$ ocorreram vieses que variaram de $0,0013$ a $0,0313$ e desvios-padrão que foram de $0,0349$ a $0,1008$.

Parâmetros dos Itens

Pode-se observar a eficácia do processo de estimação dos parâmetros dos itens na Tabela 7.4, a qual apresenta a média quadrática dos vieses-médio (MQVM) associada a cada tipo de parâmetro. Observa-se que tanto a MQVM quanto a média dos desvios-padrão (ver Tabela 7.5) apresentam-se melhor para os parâmetros c, a e b, nessa ordem. Isto pode ser devido ao fato do parâmetro b ter a mesma métrica da habilidade e das aproximações numéricas às integrais levarem em consideração pontos de quadratura no intervalo -2 a 2, o que corresponde à abrangência de habilidades apenas neste intervalo. Devemos dizer que o aumento deste intervalo resultou num aumento considerável do esforço computacional, além de não apresentar melhorias significativas dos resultados. Dois outros possíveis fatores estariam relacionados com a magnitude dos valores dos parâmetros e os valores iniciais gerados para a inicialização do processo iterativo.

Quando desejamos comparar os resultados em relação às estruturas de covariância, vemos que as estimativas para a matriz não-estruturada apresentaram-se mais consistentes. No que concerne à eficiência, a matriz uniforme com $\rho = 0,75$ foi a que apresentou melhores resultados. De um modo geral, as estimativas dos parâmetros dos itens no caso em que a matriz de covariância é não-estrutrada apresentaram o melhor resultado, o que pode ser visto na Tabela 7.6, a qual mostra as médias dos erros-médio quadráticos associadas a cada tipo de parâmetro.

Tabela 7.3: Estimativas médias e desvios-padrão das covariâncias ("correlações").

Estruturas de Covariância	
Uniforme:	$0,2731$ $(0,0113)$ $0,7589$ $(0,0076)$
Não-Estruturada:	

$$\Sigma = \begin{bmatrix} 1 & \begin{matrix}0,4646\\(0,0878)\end{matrix} & \begin{matrix}0,7464\\(0,1008)\end{matrix} & \begin{matrix}0,7427\\(0,0590)\end{matrix} & \begin{matrix}0,7292\\(0,0653)\end{matrix} \\ . & 1 & \begin{matrix}0,7605\\(0,0407)\end{matrix} & \begin{matrix}0,3820\\(0,0861)\end{matrix} & \begin{matrix}0,6717\\(0,0567)\end{matrix} \\ . & . & 1 & \begin{matrix}0,7513\\(0,0349)\end{matrix} & \begin{matrix}0,7187\\(0,0453)\end{matrix} \\ . & . & . & 1 & \begin{matrix}0,7540\\(0,0399)\end{matrix} \\ . & . & . & . & 1 \end{bmatrix}$$

Tabela 7.4: Média do quadrado (quadrática) dos vieses-médio associados a cada tipo de parâmetro do item.

	$\rho = 0,25$			$\rho = 0,75$		
	a	b	c	a	b	c
Uniforme	$0,0553$	$0,0906$	$0,0071$	$0,0606$	$0,0780$	$0,0144$
	a	b	c			
Não-Estruturada	$0,0003$	$0,0031$	$0,0006$			

Tabela 7.5: Média dos desvios-padrão associados a cada tipo de parâmetro do item.

	$\rho = 0,25$			$\rho = 0,75$		
	a	b	c	a	b	c
Uniforme	$0,2169$	$0,7515$	$0,1103$	$0,0775$	$0,1173$	$0,0356$
	a	b	c			
Não-Estruturada	$0,1472$	$0,2484$	$0,0993$			

Tabela 7.6: Média dos Erros Médio Quadrático associados a cada tipo de parâmetro do item.

	$\rho = 0,25$			$\rho = 0,75$		
	a	b	c	a	b	c
Uniforme	$0,1120$	$1,3033$	$0,0224$	$0,0672$	$0,1000$	$0,0171$
	a	b	c			
Não-Estruturada	$0,0233$	$0,0773$	$0,0127$			

7.2 Aplicação aos Dados do Vestibular 2000 - UFC

Para efeito de aplicação a dados reais, iremos analisar os dados provenientes do Vestibular 2000 da Universidade Federal do Ceará (para obter informações dos vestibulares-UFC, acessar http://sw.npd.ufc.br/ccv/). Neste vestibular, foram avaliados na primeira fase 31.016 candidatos, dos quais 30.433 compareceram aos dois dias de provas. A população de interesse do nosso estudo foi composta pelos 30.433 candidatos, da qual retiramos uma amostra aleatória de tamanho $N = 4000$. Para manter uma uniformidade com os resultados apresentados na seção anterior, iremos considerar apenas 5 provas ou disciplinas dentre as 8 aplicadas aos candidatos. As disciplinas foram: Física (15 itens, com um anulado pela Comissão Coordenadora do Vestibular - CCV), História (15 itens), Português (18 itens), Matemática (15 itens) e Biologia (15 itens). Todas as questões são de múltipla escolha com 5 alternativas e corrigidas como certo/errado. Lembramos que as respostas omitidas pelos candidatos foram corrigidas como erradas, uma vez que neste trabalho não desenvolvemos a teoria para dados incompletos. Para os valores iniciais dos parâmetros dos itens utilizamos as sugestões do capítulo 5.

Nesta análise será realizada a estimação conjunta dos parâmetros populacionais e de itens, levando-se em consideração as estruturas de covariância uniforme e não-estruturada para a distribuição das habilidades. Para quantificar qual estrutura de covariância é a mais adequada, adotaremos o valor da log-verossimilhança, o que permitirá a utilização do teste da razão de verossimilhança. A inversa da matriz Hessiana será usada para a estimação dos erros-padrão associados a cada parâmetro.

Análise dos Parâmetros Populacionais

Observando a Tabela 7.7 vemos que os resultados obtidos podem ser considerados bastante coerentes. Podemos notar, por exemplo, que a habilidade em Física está muito bem associada à habilidade em Matemática $(0,95)$ e menos associada à Português $(0,70)$. De forma análoga, o conhecimento em História está bem mais correlacionado com o conhecimento em Português $(0,81)$ do que com Matemática $(0,61)$. Para o caso em que supomos igual relação de dependência entre qualquer par de habilidades (estrutura uniforme), observamos uma alta correlação $(0,74)$. De um modo geral, observamos altas correlações que podem estar refletindo uma característica particular da população sob estudo, pois de certa forma a população de vestibulandos é homogênea uma vez que todos ou a maioria dos candidatos esforçaram-se por algum tempo para melhorar seus conhecimentos nas diversas disciplinas exigidas no concurso do Vestibular. Provavelmente os valores de algumas correlações diminuiriam se a população de estudo fosse composta por alunos matriculados em cursos regulares. Enfim, estas são apenas algumas dentre inúmeras interpretações possíveis sobre estes resultados.

Tabela 7.7: Estimativas das Covariâncias.

Estrutura de Covariância					
Uniforme: 0,7425					
Não-Estruturada:					
	Fis	**Hist**	**Por**	**Mat**	**Bio**
Fis	-	0,7132	0,6957	0,9544	0,8811
Hist	.	-	0,8114	0,6143	0,7342
Por	.	.	-	0,6266	0,7775
Mat	.	.	.	-	0,7046
Bio	.	.	.	.	-

Devemos lembrar que estes resultados são estimativas e, portanto, para afirmar por exemplo que o conhecimento em Biologia apresenta maior correlação com Português $(0,78)$ do que com História $(0,73)$, é necessário a realização de testes de hipótese. Tais testes podem ser realizados utilizando-se as distribuições assintóticas destes estimadores. Um outro fato importante, é que a partir de agora a aplicação de testes que visem à comparação dos conhecimentos ou habilidades devem levar em consideração a relação de dependência entre as habilidades envolvidas.

Para analisarmos a eficiência das estimativas podemos observar na Tabela 7.8 que os erros-padrão são bastante pequenos.

Tabela 7.8: Estimativas dos erros-padrão.

Estrutura de Covariância					
Uniforme: 0,0111					
Não-Estruturada:					
	Fis	**Hist**	**Por**	**Mat**	**Bio**
Fis	-	0,0147	0,0142	0,0010	0,0036
Hist	.	-	0,0156	0,0186	0,0147
Por	.	.	-	0,0193	0,0094
Mat	.	.	.	-	0,0063
Bio	.	.	.	.	-

Quando desejamos avaliar qual estrutura de correlação melhor se adequa para a modelagem da relação de dependência entre as habilidades, não há dúvida de que a não-estruturada é a que melhor representa esta relação, pois o teste da razão de verossimilhança apresenta-se altamente significativo (ver Tabela 7.9).

Tabela 7.9: Teste de adequação do modelo.

Modelo	$-2 \times Log - verossimilhança$	χ^2	gl	Nível descritivo
Uniforme	$397.312, 0320$	$8.308, 6276$	9	$p < 0,0001$
Não-Estruturada	$405.620, 6596$			

Análise dos Parâmetros de Itens

Alguns resultados sobre os itens que compõem cada prova e que são de grande interesse prático podem ser vistos na Tabela 7.10 para o caso da matriz de covariância não-estruturada e na Tabela 7.11 para o caso da matriz de covariância com estrutura uniforme. As interpretações a seguir serão descritas sobre a Tabela 7.10, sendo análogas para a Tabela 7.11 porque os resultados obtidos são bastante próximos.

Tabela 7.10: Estatísticas descritivas das estimativas dos parâmetros dos itens (matriz de covariância não-estruturada).

		a	b	c
Física	Mínimo	0,8499	1,1002	0,1644
	Máximo	1,1953	3,8249	0,8625
	Média	**1,0303**	**2,2158**	**0,3513**
	Desvio Padrão	0,0906	0,8814	0,1685
História	Mínimo	0,7878	-1,0710	0,0735
	Máximo	1,1787	4,5649	0,6800
	Média	**0,9666**	**0,9238**	**0,3294**
	Desvio Padrão	0,1332	1,5381	0,1808
Português	Mínimo	0,4507	-2,4279	0,0016
	Máximo	1,1157	5,5149	0,6419
	Média	**0,8541**	**1,0488**	**0,2956**
	Desvio Padrão	0,1639	2,1068	0,1783
Matemática	Mínimo	0,7900	0,4533	0,0159
	Máximo	1,1366	3,7411	0,6540
	Média	**0,9059**	**2,4560**	**0,4074**
	Desvio Padrão	0,0973	0,9838	0,1544
Biologia	Mínimo	0,7367	-0,9016	0,1458
	Máximo	1,2493	2,6759	0,9925
	Média	**1,0073**	**0,7699**	**0,3648**
	Desvio Padrão	0,1547	0,9400	0,1987

* Valores dos parâmetros na escala normal (D=1,7).

Primeiramente, podemos observar que as provas de Matemática e Física são as mais difíceis (valores médios de b iguais a $2,46$ e $2,22$, respectivamente), enquanto que as provas de Português e História apresentam um grau de dificuldade um pouco menor (b médios iguais a $1,05$ e $0,92$ respectivamente) e a prova de Biologia com o grau médio de dificuldade igual a $b = 0,77$, pode ser considerada como a menos difícil, isto porque estes valores estão todos acima do valor médio da escala, $b = 0,0$. Em relação ao parâmetro de discriminação, podemos dizer que as provas conseguem atingir o objetivo de discriminar bem os candidatos do vestibular, pois em média estes valores variam de $0,85$ a $1,03$ e que na escala normal são considerados bons valores. Quanto ao parâmetro de acerto casual c, devemos chamar a atenção para o valor máximo estimado na prova de Biologia $(0,99)$, ou seja, um indivíduo de baixa habilidade quase que certamente responderá corretamente este item. Valores aberrantes, como este, para os parâmetros do item podem ser considerados como uma indicação de que o item não adequa-se bem ao modelo proposto, podendo estar violando, por exemplo, a suposição de unidimensionalidade do traço latente (ver Cap. 2). Esta situação requer uma análise detalhada dos itens envolvidos nas provas e que não foi realizada neste trabalho por não contemplar os objetivos principais.

Tabela 7.11: Estatísticas descritivas das estimativas dos parâmetros dos itens (matriz de covariância uniforme).

		a	b	c
Física	Mínimo	0,8502	1,0746	0,0006
	Máximo	1,2179	3,8250	0,5093
	Média	**1,0335**	**2,2136**	**0,3009**
	Desvio Padrão	0,0947	0,8854	0,1244
História	Mínimo	0,7784	-1,0673	0,0735
	Máximo	1,1785	4,5649	0,6547
	Média	**0,9641**	**0,9341**	**0,3105**
	Desvio Padrão	0,1320	1,5348	0,1536
Português	Mínimo	0,4813	-2,4288	0,0026
	Máximo	1,1107	5,5149	0,8248
	Média	**0,8484**	**1,0510**	**0,3075**
	Desvio Padrão	0,1595	2,1064	0,2064
Matemática	Mínimo	0,7905	0,4944	0,2220
	Máximo	1,1519	3,7414	0,7241
	Média	**0,9046**	**2,4540**	**0,4425**
	Desvio Padrão	0,0970	0,9847	0,1275
Biologia	Mínimo	0,7387	-0,9035	0,1476
	Máximo	1,2537	2,6758	0,9695
	Média	**1,0061**	**0,7667**	**0,3916**
	Desvio Padrão	0,1582	0,9410	0,2342

* Valores dos parâmetros na escala normal (D=1,7).

Na Tabela 7.12 podemos observar a eficiência dos parâmetros de itens estimados, uma vez que as médias dos erros-padrão são bem pequenas.

Tabela 7.12: Média dos erros-padrão associados a cada tipo de parâmetro de item.

Modelo	a	b	c
Uniforme	$0,0321$	$0,0631$	$0,0052$
Não-Estruturada	$0,0500$	$0,0915$	$0,0053$

É importante deixarmos claro que os resultados aqui apresentados foram obtidos a partir da utilização de um programa computacional que ainda está em fase de desenvolvimento. Em algumas situações o critério de convergência não foi alcançado.

Capítulo 8

Considerações Finais

Para concluir o nosso trabalho, comentaremos algumas dificuldades encontradas que não foram alvo de estudos mais detalhados, ficando portanto como sugestões para futuros trabalhos na Teoria da Resposta ao Item. Outras sugestões foram motivadas a partir de observações práticas.

Sem dúvida, uma das grandes dificuldades da implementação da TRI está associada com a estimação dos parâmetros dos itens, em particular nos modelos que envolvem o parâmetro c de acerto casual. Para que este parâmetro possa ser bem estimado, é necessário um bom número de respondentes com baixa habilidade. Este fato nem sempre ocorre na prática, em particular em situações onde há distribuições assimétricas negativa para as habilidades. Diante desta dificuldade, sugere-se o estudo de *Modelos Multivariados da TRI* para *Distribuições Assimétricas*, o estudo de métodos para *Pequenas Amostras* e de *Métodos Não-Paramétricos*. Problemas de instabilidade também podem ocorrer para os outros parâmetros envolvidos no modelo.

Outros problemas que podem surgir na implementação computacional dos métodos de estimação estão relacionados com os valores iniciais dos parâmetros e a ocorrência de Matriz Hessiana não-definida nos processos iterativos [ver Seber, 1989], bem como para a matriz de covariâncias. Com relação aos "chutes iniciais" para os parâmetros dos itens, existem algumas sugestões para o caso univariado que foram utilizadas neste trabalho, mas que devem ser melhor investigadas para o caso multivariado. Para os parâmetros de covariância ou correlação os valores iniciais devem levar em consideração as diferentes estruturas. Com relação à ocorrência de Matriz Hessiana não-definida em iterações do método Newton-Raphson, sugere-se o estudo de aproximações a esta matriz (ver método de Quasi-Newton em Seber, 1989) e métodos para estruturas de covariâncias que apresentam-se não-definidas também devem ser estudados. Uma outra sugestão está relacionada com o estudo de estruturas de covariância intermediárias, ou seja, estruturas que assumem formas específicas de acordo com o interesse do investigador [Singer & Andrade (2000) apresentam algumas estruturas].

Como extensões do Modelo Multivariado, pode-se propor *Modelos Multidimensionais Multivariados* (ver Nojosa, (2001), para o caso univariado) e *Modelos Longitudinais Multivariados* (ver Tavares, (2001), para o caso univariado). Uma outra extensão seria a composição dessas duas extensões gerando *Modelos Multidimensionas Longitudinais Multivariados*. Todos os modelos referidos acima podem ser generalizados para situações onde há duas ou mais populações de interesse.

Uma outra proposta de estudo seria o desenvolvimento dos modelos referidos acima para os casos de *dados incompletos*, ou seja, para os casos em que não temos todas as respostas de todos os indivíduos para todas as disciplinas de interesse.

Neste trabalho exploramos a aplicação da TRI em Avaliações Educacionais. No entanto, devemos ressaltar que a *aplicação da TRI não se restringe a esta área*, podendo, por exemplo, ser aplicada à Psicologia, Medicina, Marketing e em outras áreas ainda não exploradas. Algumas referências recentes seriam: DeRoos & Allen-Meares (1998) em psiquiatria, Tennant et al. (1996) e Granger et al. (1998) em reabilitação médica e Marcoulides (1998) que descreve o modelo Rasch ou modelo logístico de um parâmetro como um dos métodos modernos mais importantes para a pesquisa na área de negócios. Outro ponto importante e que deve ser ressaltado diz respeito à necessidade de *integração com profissionais das mais diversas áreas* para que não haja um isolamento da Teoria por falta de interpretações práticas e inovadoras dos resultados obtidos pela TRI.

Apêndice A

Desenvolvimentos Complementares: Matriz Hessiana

Os métodos iterativos para o cálculo das Estimativas de Máxima Verossimilhança-EMV são bastante utilizados na prática e, como vimos, as equações de estimação para os parâmetros dos itens e populacionais não apresentam soluções explícitas, o que nos leva à utilização de algum processo iterativo para obter as estimativas desejadas.

Expandindo as equações de verossimilhança (função escore em $\widehat{\boldsymbol{\theta}}$) em série multivariada de Taylor até primeira ordem ao redor de um ponto qualquer $\boldsymbol{\theta}$ pertencente a uma vizinhança de $\widehat{\boldsymbol{\theta}}$, obtém-se uma relação aproximada entre a EMV, a função escore e a informação observada (derivadas segundas da log-verossimilhança) avaliadas no ponto $\boldsymbol{\theta}$ próximo de $\widehat{\boldsymbol{\theta}}$, assim uma nova estimativa $\boldsymbol{\theta}^{(t+1)}$ pode ser obtida a partir de uma anterior $\boldsymbol{\theta}^{(t)}$ através de

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left[\mathbf{H}\left(\boldsymbol{\theta}^{(t)}\right)\right]^{-1} \mathbf{f}\left(\boldsymbol{\theta}^{(t)}\right),$$

onde $\mathbf{H}\left(\boldsymbol{\theta}^{(t)}\right)$ é a matriz de derivadas segundas da log-verossimilhança e $\mathbf{f}\left(\boldsymbol{\theta}^{(t)}\right)$ o vetor de derivadas primeira na $t - ésima$ iteração.

O processo é então repetido até que algum critério de parada seja satisfeito. Este procedimento é conhecido como *Algoritmo* ou *Método de Newton-Raphson* e uma forma alternativa assintótica da expressão acima é obtida ao substituir as derivadas segundas da log-verossimilhança pelos seus valores esperados (informação esperada), o que resulta no processo conhecido como *Método Escore de Fisher* [Rao, 1973]. Uma vez que o cálculo das esperanças das derivadas segundas não é uma tarefa simples, pois a variável U_{jik} surge no denominador de $g_j^*(\boldsymbol{\theta}_j)$, iremos utilizar neste trabalho o processo iterativo de

Newton-Raphson.

A seguir apresentamos as derivadas segundas necessárias para a aplicação do procedimento iterativo. As derivadas segundas para os parâmetros dos itens, populacionais, conjunta e para a moda da distribuição das habilidades serão apresentadas separadamente, de acordo com o que foi exposto nos Capítulos 3, 4, 5 e 6.

A.1 Parâmetros dos Itens

Inicialmente devemos obter a derivada segunda de $\log L(\boldsymbol{\beta}, \boldsymbol{\eta})$ em relação a $\boldsymbol{\beta}_{ik}$ e $\boldsymbol{\beta}_{lk}$. Assim, de (3.5) obtemos

$$\frac{\partial^2 \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik} \partial \boldsymbol{\beta}_{lk}^{'}} = \frac{\partial}{\partial \boldsymbol{\beta}_{ik}} \left[\sum_{j=1}^{s} r_j \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{lk}} \right]^{'}$$

$$= \sum_{j=1}^{s} r_j \left[\frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik} \partial \boldsymbol{\beta}_{lk}^{'}} - \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}) / \partial \boldsymbol{\beta}_{ik}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \right) \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}) / \partial \boldsymbol{\beta}_{lk}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \right)^{'} \right], \tag{A.1}$$

onde a segunda parcela de (A.1) é obtida de (3.11). Agora vejamos na primeira parcela o cálculo da $\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik} \partial \boldsymbol{\beta}_{lk}^{'}}$.

Para simplificar, consideremos dois casos: $\begin{cases} \text{Caso 1: } i = l \\ \text{Caso 2: } i \neq l \end{cases}$.

Caso 1: $i = l$

De (3.11) temos que

$$\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\beta}_{ik}^{'}} = \frac{\partial}{\partial\boldsymbol{\beta}_{ik}}\left\{\int_{I\!R^K}\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}\right\}^{'}$$
$$= \int_{I\!R^K}\left\{\frac{\partial}{\partial\boldsymbol{\beta}_{ik}}\left[\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}\right]P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})+\right.$$
$$\left.\left[\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\right]\left[\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{ik}}\right]^{'}\right\}g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}. \quad \text{(A.2)}$$

Vejamos inicialmente a primeira parcela de (A.2). Façamos $V_{jik} = \frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}$, de forma que

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}_{ik}}\left[\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}\right] &= \frac{\partial}{\partial\boldsymbol{\beta}_{ik}}\left[V_{jik}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}\right]\\ &= \frac{\partial V_{jik}}{\partial\boldsymbol{\beta}_{ik}}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}+V_{jik}\left(\frac{\partial^2 P_{ik}}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\beta}_{ik}^{'}}\right). \quad \text{(A.2.1)}\end{aligned}$$

Mas note que

$$\begin{aligned}\frac{\partial V_{jik}}{\partial\boldsymbol{\beta}_{ik}} &= \frac{\partial}{\partial\boldsymbol{\beta}_{ik}}\left[\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\right]=\\ &= \frac{1}{(P_{ik}Q_{ik})^2}\left[-\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}P_{ik}Q_{ik}-(U_{jik}-P_{ik})\frac{\partial P_{ik}Q_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right].\end{aligned}$$

Sabendo que $\frac{\partial P_{ik}Q_{ik}}{\partial\boldsymbol{\beta}_{ik}} = (1-2P_{ik})\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}$, temos

$$\begin{aligned}\frac{\partial V_{jik}}{\partial\boldsymbol{\beta}_{ik}} &= \frac{-1}{(P_{ik}Q_{ik})^2}\left[\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}P_{ik}Q_{ik}+(U_{jik}-P_{ik})(1-2P_{ik})\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right]\\ &= \frac{-1}{(P_{ik}Q_{ik})^2}\left[P_{ik}Q_{ik}+(U_{jik}-P_{ik})(1-2P_{ik})\right]\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}. \quad \text{(A.2.2)}\end{aligned}$$

Resolvendo a expressão entre colchetes, temos

$$P_{ik}Q_{ik}+(U_{jik}-P_{ik})(1-2P_{ik}) = P_{ik}Q_{ik}+U_{jik}-2P_{ik}U_{jik}-P_{ik}+2P_{ik}^2$$
$$= P_{ik}Q_{ik}-P_{ik}+P_{ik}^2+(U_{jik}-2P_{ik}U_{jik}+P_{ik}^2)$$

$= P_{ik}Q_{ik} - P_{ik}\left(1 - P_{ik}\right) + (U_{jik} - P_{ik})^2$
$= P_{ik}Q_{ik} - P_{ik}Q_{ik} + (U_{jik} - P_{ik})^2$
$= (U_{jik} - P_{ik})^2.$

Em que $(U_{jik} - 2P_{ik}U_{jik} + P_{ik}^2) = (U_{jik} - P_{ik})^2$ pelo fato de $U_{jik} = U_{jik}^2$. Agora, substituindo o resultado anterior na expressão (A.2.2), obtemos

$$\begin{aligned}\frac{\partial V_{jik}}{\partial \boldsymbol{\beta}_{ik}} &= \frac{-1}{\left(P_{ik}Q_{ik}\right)^2}\left[(U_{jik} - P_{ik})^2\right]\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}} = \\ &= -V_{jik}^2\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}. \qquad \text{(A.2.3)}\end{aligned}$$

Substituindo (A.2.3) em (A.2.1), segue que

$$\frac{\partial}{\partial \boldsymbol{\beta}_{ik}}\left[V_{jik}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)^{'}\right] = -V_{jik}^2\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)^{'} + V_{jik}\left(\frac{\partial^2 P_{ik}}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}_{ik}^{'}}\right). \qquad \text{(A.2.4)}$$

Agora vejamos a segunda parcela de (A.2). Observando-se o desenvolvimento de (3.7) a (3.11), verificamos que

$$\begin{aligned}\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{ik}} &= \left\{\frac{\partial}{\partial \boldsymbol{\beta}_{ik}}\log P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta}) \\ &= \left\{\frac{(U_{jik} - P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta}) \\ &= V_{jik}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta}). \qquad \text{(A.2.5)}\end{aligned}$$

Finalmente, podemos substituir os resultados (A.2.4) e (A.2.5) em (A.2), assim segue

$$\begin{aligned}
\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\beta}_{ik}^{'}} &= \int_{I\!\!R^K}\left\{\left[V_{jik}\left(\frac{\partial^2 P_{ik}}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\beta}_{ik}^{'}}\right) - V_{jik}^2\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}\right]+\right.\\
&\left.+\left[V_{jik}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\right]\left[V_{jik}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\right]^{'}\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}\\
&= \int_{I\!\!R^K}\left\{V_{jik}\left(\frac{\partial^2 P_{ik}}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\beta}_{ik}^{'}}\right) - V_{jik}^2\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}+\right.\\
&\left.+V_{jik}^2\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)^{'}\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}\\
&= \int_{I\!\!R^K}V_{jik}\left(\frac{\partial^2 P_{ik}}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\beta}_{ik}^{'}}\right)P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}. \qquad \text{(A.3)}
\end{aligned}$$

Restando agora o cálculo das derivadas segundas específicas, as quais para o caso do ML3 são dadas por

$$\begin{aligned}
\frac{\partial^2 P_{ik}}{\partial a_{ik}^2} &= D^2\left(\theta_k - b_{ik}\right)^2\left(1-c_{ik}\right)\left[1-2P_{ik}^*\right]P_{ik}^*Q_{ik}^*, && \text{(A.4)}\\
\frac{\partial^2 P_{ik}}{\partial a_{ik}\partial b_{ik}} &= -D\left(1-c_{ik}\right)\left\{1+Da_{ik}\left(\theta_k-b_{ik}\right)\left[1-2P_{ik}^*\right]\right\}P_{ik}^*Q_{ik}^*, && \text{(A.5)}\\
\frac{\partial^2 P_{ik}}{\partial a_{ik}\partial c_{ik}} &= -D\left(\theta_k-b_{ik}\right)P_{ik}^*Q_{ik}^*, && \text{(A.6)}\\
\frac{\partial^2 P_{ji}}{\partial b_{ik}^2} &= D^2a_{ik}^2\left(1-c_{ik}\right)\left[1-2P_{ik}^*\right]P_{ik}^*Q_{ik}^*, && \text{(A.7)}\\
\frac{\partial^2 P_{ik}}{\partial b_{ik}\partial c_{ik}} &= Da_{ik}P_{ik}^*Q_{ik}^*, && \text{(A.8)}\\
\frac{\partial^2 P_{ik}}{\partial c_{ik}^2} &= 0. && \text{(A.9)}
\end{aligned}$$

Para simplificar usaremos as seguintes notações:

$$\mathbf{h}_{ik} = \left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right) = \left(\begin{array}{c} D(1-c_{ik})(\theta_k - b_{ik}) \\ -Da_{ik}(1-c_{ik})P^*_{ik}Q^*_{ik} \\ \frac{1}{P^*_{ik}} \end{array}\right) P^*_{ik}Q^*_{ik},$$

$$\begin{aligned}
\mathbf{H}_{iik} &= \left(\frac{\partial^2 P_{ik}}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}^{'}_{ik}}\right) \\
&= \left(\begin{array}{ccc} D^2\left(\theta_k - b_{ik}\right)^2\left(1-c_{ik}\right)\left[1-2P^*_{ik}\right] & . & . \\ -D\left(1-c_{ik}\right)\left\{1+Da_{ik}\left(\theta_k - b_{ik}\right)\left[1-2P^*_{ik}\right]\right\} & D^2a^2_{ik}\left(1-c_{ik}\right)\left[1-2P^*_{ik}\right] & . \\ -D\left(\theta_k - b_{ik}\right) & Da_{ik} & 0 \end{array}\right) P^*_{ik}Q^*_{ik}.
\end{aligned}$$

De (A.3) para usar em (A.1) denotaremos

$$\begin{aligned}
\mathbf{H}_{ii(j)} &= \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}^{'}_{ik}} \\
&= \int_{I\!R^K}\left\{V_{jik}\left(\frac{\partial^2 P_{ik}}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}^{'}_{ik}}\right)\right\}\frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}d\boldsymbol{\theta} \\
&= \int_{I\!R^K}\left\{V_{jik}\mathbf{H}_{iik}\right\}g^*_j(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}.
\end{aligned}$$

Caso 2: $i \neq l$

De forma similar ao caso anterior, temos

$$\begin{aligned}
\mathbf{H}_{il(j)} &= \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}^{'}_{lk}} \\
&= \int_{I\!R^K}\left\{\left[V_{jlk}\left(\frac{\partial P_{lk}}{\partial \boldsymbol{\beta}_{lk}}\right)\right]\left[V_{jik}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\right]^{'}\right\}g^*_j(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&= \int_{I\!R^K}\left\{\left[V_{jlk}\mathbf{h}_{lk}\right]\left[V_{jik}\mathbf{h}^{'}_{ik}\right]\right\}g^*_j(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta},
\end{aligned}$$

pois observe que na primeira parcela de (A.2) teremos

$$\frac{\partial}{\partial \boldsymbol{\beta}_{ik}}\left[\frac{(U_{jlk}-P_{lk})}{P_{lk}Q_{lk}}\left(\frac{\partial P_{lk}}{\partial \boldsymbol{\beta}_{lk}}\right)^{'}\right]=\frac{\partial}{\partial \boldsymbol{\beta}_{ik}}\left[\mathbf{f}\left(\boldsymbol{\beta}_{lk}\right)\right]=\mathbf{0},$$

restando apenas a segunda parcela, o que justifica o resultado obtido para $\mathbf{H}_{il(j)}$.

Para a segunda parcela em (A.1) usamos notação análoga, de forma que

$$\begin{aligned}
\mathbf{h}_{i(j)} &= \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})/\partial \boldsymbol{\beta}_{ik}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}\\
&= \int_{I\!R^K}\left\{\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}}\right)\right\}\frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}d\boldsymbol{\theta}\\
&= \int_{I\!R^K} V_{jik}\mathbf{h}_{ik}g_j^*(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}.
\end{aligned}$$

Assim, retornando a (A.1), obtemos

$$\begin{aligned}
\mathbf{H}\left(\boldsymbol{\beta}_{ik},\boldsymbol{\beta}_{lk}\right) &= \frac{\partial^2 \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik}\partial \boldsymbol{\beta}_{lk}^{'}}\\
&= \sum_{j=1}^{N}\left[\mathbf{H}_{il(j)}-\mathbf{h}_{i(j)}\mathbf{h}_{l(j)}^{'}\right]. \qquad \text{(A.10)}
\end{aligned}$$

Ainda,

$$\mathbf{f}_1\left(\boldsymbol{\beta}_{ik}\right)=\frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial \beta_{ik}}=\sum_{j=1}^{s} r_j\int_{I\!R^K} V_{jik}\mathbf{h}_{ik}g_j^*(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta},$$

$$\mathbf{f}_{\mathbf{I}}\left(\boldsymbol{\beta}\right)=\begin{pmatrix}\mathbf{f}_1\left(\boldsymbol{\beta}_{11}\right)\\ \ldots\\ \mathbf{f}_1\left(\boldsymbol{\beta}_{n_1 1}\right)\\ \ldots\\ \mathbf{f}_1\left(\boldsymbol{\beta}_{1k}\right)\\ \ldots\\ \mathbf{f}_1\left(\boldsymbol{\beta}_{n_k k}\right)\end{pmatrix} \text{ e}$$

$$
\mathbf{H_I}(\boldsymbol{\beta}) = \begin{pmatrix}
\begin{matrix} \mathbf{H}(\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{11}) & ... & \mathbf{H}(\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{n_1 1}) \\ ... & ... & ... \\ \mathbf{H}(\boldsymbol{\beta}_{n_1 1}, \boldsymbol{\beta}_{11}) & ... & \mathbf{H}(\boldsymbol{\beta}_{n_1 1}, \boldsymbol{\beta}_{n_1 1}) \end{matrix} & & \\
& ... & \\
& & \begin{matrix} \mathbf{H}(\boldsymbol{\beta}_{1k}, \boldsymbol{\beta}_{1k}) & ... & \mathbf{H}(\boldsymbol{\beta}_{1k}, \boldsymbol{\beta}_{n_k k}) \\ ... & ... & ... \\ \mathbf{H}(\boldsymbol{\beta}_{n_k k}, \boldsymbol{\beta}_{1k}) & ... & \mathbf{H}(\boldsymbol{\beta}_{n_k k}, \boldsymbol{\beta}_{n_k k}) \end{matrix}
\end{pmatrix}.
$$

Assim, temos que se $\widehat{\boldsymbol{\beta}}^{(t)}$ é uma estimativa de $\boldsymbol{\beta}$ na iteração t, então na iteração $t+1$ teremos

$$
\widehat{\boldsymbol{\beta}}^{(t+1)} = \widehat{\boldsymbol{\beta}}^{(t)} - \left[\mathbf{H_I}\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)\right]^{-1} \mathbf{f_I}\left(\widehat{\boldsymbol{\beta}}^{(t)}\right).
$$

A.2 Parâmetros Populacionais

A partir da expressão (4.1), podemos obter a derivada segunda da log-verossimilhança, como segue:

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} &= \frac{\partial}{\partial \boldsymbol{\eta}} \left\{ \sum_{j=1}^{s} r_j \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right\}' \\
&= \sum_{j=1}^{s} r_j \left\{ \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \right. \\
&\quad \left. - \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \right) \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} \right)' \right\}, \qquad \text{(A.11)}
\end{aligned}
$$

onde o segundo termo pode ser obtido por (4.3). Agora vejamos a primeira parcela:

$$\begin{aligned}
\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'} &= \frac{\partial}{\partial\boldsymbol{\eta}}\left\{\int_{I\!R^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})\left(\frac{\partial}{\partial\boldsymbol{\eta}}\log g(\boldsymbol{\theta}|\boldsymbol{\eta})\right)d\boldsymbol{\theta}\right\}' \\
&= \int_{I\!R^K}\left\{\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right. \\
&\quad \left. + \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)\left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)'\right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta},
\end{aligned}$$

faltando, portanto, apenas as derivadas segundas de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta})$, que para o caso Normal, fica

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\mu_k\partial\mu_l'} = -v_{k,l}, \quad k,l = 1,...,K,$$

que, em notação matricial, resulta em

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}'} = -\mathbf{V}. \tag{A.12}$$

Ainda,

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\mu}\partial\sigma_{kl}} = \left[\frac{\partial\mathbf{V}}{\partial\sigma_{kl}}\right](\boldsymbol{\theta}-\boldsymbol{\mu}) \tag{A.13}$$

e

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\sigma_{kl}\partial\sigma_{k'l'}} = -\frac{1}{2}\frac{\partial^2 \log|\Sigma|}{\partial\sigma_{kl}\partial\sigma_{k'l'}} - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\left[\frac{\partial^2\mathbf{V}}{\partial\sigma_{kl}\partial\sigma_{k'l'}}\right](\boldsymbol{\theta}-\boldsymbol{\mu}). \tag{A.14}$$

Usando a notação

$$\mathbf{f_P}(\boldsymbol{\eta}) = \frac{\partial \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}, \qquad \mathbf{H_P}(\boldsymbol{\eta}) = \frac{\partial^2 \log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}$$

e se $\hat{\boldsymbol{\eta}}^{(t)}$ é uma estimativa de $\boldsymbol{\eta}$ na iteração t, então na iteração $t+1$ teremos

$$\hat{\boldsymbol{\eta}}^{(t+1)} = \hat{\boldsymbol{\eta}}^{(t)} - \left[\mathbf{H_P}\left(\hat{\boldsymbol{\eta}}^{(t)}\right)\right]^{-1} \mathbf{f_P}\left(\hat{\boldsymbol{\eta}}^{(t)}\right). \tag{A.15}$$

A seguir, veremos como ficam as derivadas segundas para o caso em que a matriz de variância-covariância é não-estruturada e posteriormente para algumas estruturas particulares.

A.2.1 Matriz Não-Estruturada

A partir de (A.13) temos que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \sigma_{kl}} = \left[-\Sigma^{-1}\Delta_{kl}\Sigma^{-1}\right](\boldsymbol{\theta} - \boldsymbol{\mu}),$$

onde Δ_{kl} é uma matriz $K \times K$ cujo elemento kl e lk é igual a 1 e os demais são iguais a zero. Para obter a expressão (A.14) precisamos de

$$\begin{aligned}
\frac{\partial^2 \log|\Sigma|}{\partial \sigma_{kl}\partial \sigma_{k'l'}} &= \frac{\partial}{\partial \sigma_{kl}}\left\{\frac{\partial \log|\Sigma|}{\partial \sigma_{k'l'}}\right\}' \\
&= \begin{cases} \frac{\partial}{\partial \sigma_{kl}}\sigma^{k'k'}; & k' = l' \\ 2\frac{\partial}{\partial \sigma_{kl}}\sigma^{k'l'}; & k' \neq l', \end{cases}
\end{aligned}$$

e como desconhecemos a forma dos elementos de Σ^{-1} os quais são funções dos elementos de Σ, não conseguimos encontrar uma forma fechada para esta derivada. Uma possível solução seria a utilização de aproximações numéricas. Precisamos também de

$$
\begin{aligned}
\frac{\partial^2 \mathbf{V}}{\partial \sigma_{kl} \partial \sigma_{k'l'}} &= \frac{\partial}{\partial \sigma_{kl}} \left\{ \frac{\partial \mathbf{V}}{\partial \sigma_{_{k'l'}}} \right\} = \frac{\partial}{\partial \sigma_{kl}} \left\{ \frac{\partial \Sigma^{-1}}{\partial \sigma_{_{k'l'}}} \right\} \\
&= \frac{\partial}{\partial \sigma_{kl}} \left\{ -\Sigma^{-1} \Delta_{_{k'l'}} \Sigma^{-1} \right\} \\
&= \left[\frac{\partial}{\partial \sigma_{kl}} \left(-\Sigma^{-1} \Delta_{_{k'l'}} \right) \right] \Sigma^{-1} + \left(-\Sigma^{-1} \Delta_{_{k'l'}} \right) \left(\frac{\partial \Sigma^{-1}}{\partial \sigma_{kl}} \right) \\
&= \left[-\left(\frac{\partial \Sigma^{-1}}{\partial \sigma_{kl}} \right) \Delta_{_{k'l'}} \Sigma^{-1} - \Sigma^{-1} \left(\frac{\partial \Delta_{_{k'l'}}}{\partial \sigma_{kl}} \right) \Sigma^{-1} \right] - \left(\Sigma^{-1} \Delta_{_{k'l'}} \right) \left(-\Sigma^{-1} \Delta_{_{kl}} \Sigma^{-1} \right) \\
&= \left(\Sigma^{-1} \Delta_{_{kl}} \Sigma^{-1} \right) \Delta_{_{k'l'}} \Sigma^{-1} + \Sigma^{-1} \Delta_{_{k'l'}} \Sigma^{-1} \Delta_{_{kl}} \Sigma^{-1} \\
&= \Sigma^{-1} \Delta_{_{kl}} \Sigma^{-1} \Delta_{_{k'l'}} \Sigma^{-1} + \Sigma^{-1} \Delta_{_{k'l'}} \Sigma^{-1} \Delta_{_{kl}} \Sigma^{-1} \\
&= \Sigma^{-1} \left[\Delta_{_{kl}} \Sigma^{-1} \Delta_{_{k'l'}} + \Delta_{_{k'l'}} \Sigma^{-1} \Delta_{_{kl}} \right] \Sigma^{-1},
\end{aligned}
$$

pois $\left(\frac{\partial \Delta_{_{k'l'}}}{\partial \sigma_{kl}} \right) = \mathbf{0}$.

Desta forma, para chegar às estimativas procuradas basta substituir as derivadas segundas em (A.11) e aplicar em (A.15). Veremos a seguir as derivadas segundas para a estrutura matricial uniforme.

A.2.2 Matriz Uniforme

Para essa estrutura as derivadas segundas são dadas por

$$
\begin{aligned}
\frac{\partial^2 \log |\Sigma|}{\partial \rho^2} &= -(K-1) \left[\frac{1}{\alpha^2} - \frac{(K-1)}{\beta^2} \right] \\
\mathbf{V}_\rho^{(2)} &\equiv \frac{\partial^2 \mathbf{V}}{\partial \rho^2} = \frac{2}{\sigma^2 \alpha^3} \mathbf{I}_K + \frac{2}{\sigma^2 (\alpha\beta)^3} \left[(\delta - (K-1)\rho) \alpha\beta - \rho\delta^2 \right] \mathbf{J}_K \\
\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \rho} &= \left[\frac{\partial \mathbf{V}}{\partial \rho} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}) = \mathbf{V}_\rho^{(1)} (\boldsymbol{\theta} - \boldsymbol{\mu}),
\end{aligned}
$$

e com relação a σ^2 temos

$$
\frac{\partial^2 \log |\Sigma|}{\partial (\sigma^2)^2} = -\frac{K}{\sigma^4}, \qquad \frac{\partial^2 \log |\Sigma|}{\partial \sigma^2 \partial \rho} = 0,
$$

$$\frac{\partial^2 \mathbf{V}}{\partial \left(\sigma^2\right)^2} = \frac{2}{\sigma^4}\mathbf{V}, \qquad \frac{\partial^2 \mathbf{V}}{\partial \sigma^2 \partial \rho} = -\frac{1}{\sigma^2}\mathbf{V}_\rho^{(1)}$$

e

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \sigma^2} = \left[\frac{\partial \mathbf{V}}{\partial \sigma^2}\right](\boldsymbol{\theta} - \boldsymbol{\mu}) = -\frac{1}{\sigma^2}\mathbf{V}\left(\boldsymbol{\theta} - \boldsymbol{\mu}\right).$$

A.3 Parâmetros dos Itens e Parâmetros Populacionais

De acordo com o que foi descrito em relação à métrica dos parâmetros, não precisaremos estimar o vetor de médias $\boldsymbol{\mu}$, pois para fixar a métrica estabelecemos que $\boldsymbol{\mu} = \mathbf{0}$, de forma análoga não precisaremos estimar as variâncias, pois $\sigma_{11} = \sigma_{22} = ... = \sigma_{KK} = 1$ e portanto é necessário apenas estimar as covariâncias, que neste caso é equivalente às correlações entre as K habilidades. Desta forma temos que

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \sigma_{kl}} = -\frac{1}{2}\frac{\partial}{\partial \sigma_{kl}}\left(\log |\Sigma|\right) - \frac{1}{2}\boldsymbol{\theta}'\left[\frac{\partial \mathbf{V}}{\partial \sigma_{kl}}\right]\boldsymbol{\theta}$$

e

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \sigma_{kl} \partial \sigma_{k'l'}} = -\frac{1}{2}\frac{\partial \log |\Sigma|}{\partial \sigma_{kl} \partial \sigma_{k'l'}} - \frac{1}{2}\boldsymbol{\theta}'\left[\frac{\partial^2 \mathbf{V}}{\partial \sigma_{kl} \partial \sigma_{k'l'}}\right]\boldsymbol{\theta}.$$

A partir da expressão (4.1), temos que a derivada segunda cruzada entre $\boldsymbol{\beta}_{ik}$ e $\boldsymbol{\eta}$ é dada por

$$\frac{\partial \log L(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta}_{ik} \partial \boldsymbol{\eta}'} = \sum_{j=1}^{s} r_j \left\{ \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})/\partial \boldsymbol{\beta}_{ik} \partial \boldsymbol{\eta}'}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})} - \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})/\partial \boldsymbol{\beta}_{ik}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}\right)\left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})/\partial \boldsymbol{\eta}}{P(\mathbf{U}_{j..}|\boldsymbol{\beta}, \boldsymbol{\eta})}\right)' \right\} \tag{A.16}$$

em que a segunda parcela é obtida pelas expressões (3.11) e (4.3), respectivamente.

Agora, em relação à primeira parcela, temos de (4.3) que

$$\begin{aligned}
\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\eta}'} &= \frac{\partial}{\partial\boldsymbol{\beta}_{ik}}\left\{\int_{I\!R^K} P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})\left(\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)d\boldsymbol{\theta}\right\}' \\
&= \int_{I\!R^K}\left\{\left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)'\right\}g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&= \int_{I\!R^K}\left\{\left(\frac{\partial\log P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)'\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\
&= \int_{I\!R^K}\left\{\frac{(U_{jik}-P_{ik})}{P_{ik}Q_{ik}}\left(\frac{\partial P_{ik}}{\partial\boldsymbol{\beta}_{ik}}\right)\left(\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)'\right\}P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}.
\end{aligned}$$

Assim, obtemos como resultado para a primeira parcela de (A.16)

$$\begin{aligned}
\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})/\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\eta}'}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})} &= \int_{I\!R^K}\left\{V_{jik}\mathbf{h}_{ik}\left(\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)'\right\}\frac{P(\mathbf{U}_{j..}|\boldsymbol{\theta},\boldsymbol{\beta})g(\boldsymbol{\theta}|\boldsymbol{\eta})}{P(\mathbf{U}_{j..}|\boldsymbol{\beta},\boldsymbol{\eta})}d\boldsymbol{\theta} \\
&= \int_{I\!R^K}\left\{V_{jik}\mathbf{h}_{ik}\left(\frac{\partial\log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right)'\right\}g_j^*(\boldsymbol{\theta})d\boldsymbol{\theta}. \qquad \text{(A.17)}
\end{aligned}$$

Portanto,

$$\mathbf{H}_{ikP}(\boldsymbol{\beta}_{ik},\boldsymbol{\eta}) = \frac{\partial\log L(\boldsymbol{\beta},\boldsymbol{\eta})}{\partial\boldsymbol{\beta}_{ik}\partial\boldsymbol{\eta}'},\ \ \mathbf{H}_{IP}(\boldsymbol{\beta},\boldsymbol{\eta}) = \begin{pmatrix} \mathbf{H}_{11P}(\boldsymbol{\beta}_{11},\boldsymbol{\eta}) \\ \dots \\ \mathbf{H}_{n_11P}(\boldsymbol{\beta}_{n_11},\boldsymbol{\eta}) \\ \dots \\ \mathbf{H}_{1KP}(\boldsymbol{\beta}_{1K},\boldsymbol{\eta}) \\ \dots \\ \mathbf{H}_{n_KKP}(\boldsymbol{\beta}_{n_KK},\boldsymbol{\eta}) \end{pmatrix}.$$

Sejam $\boldsymbol{\psi} = (\boldsymbol{\beta}',\boldsymbol{\eta}')'$ e

$$\mathbf{f}(\boldsymbol{\psi}) = \begin{pmatrix} \mathbf{f}_I(\boldsymbol{\beta}) \\ \mathbf{f}_P(\boldsymbol{\eta}) \end{pmatrix},\ \ \mathbf{H}(\boldsymbol{\psi}) = \begin{pmatrix} \mathbf{H}_I(\boldsymbol{\beta}) & \mathbf{H}_{IP}(\boldsymbol{\beta},\boldsymbol{\eta}) \\ \mathbf{H}_{IP}(\boldsymbol{\beta},\boldsymbol{\eta})' & \mathbf{H}_P(\boldsymbol{\eta}) \end{pmatrix}.$$

Se $\hat{\boldsymbol{\psi}}^{(t)}$ é uma estimativa de $\boldsymbol{\psi}$ na iteração t, então na iteração $t+1$ teremos que

$$\hat{\boldsymbol{\psi}}^{(t+1)} = \hat{\boldsymbol{\psi}}^{(t)} - \left[\mathbf{H}\left(\hat{\boldsymbol{\psi}}^{(t)}\right)\right]^{-1} \mathbf{f}\left(\hat{\boldsymbol{\psi}}^{(t)}\right).$$

A.3.1 Expressões para o Algoritmo EM

As expressões para aplicação do algoritmo EM são obtidas ao se considerar fixas as quantidades f_{ik}, r_{ik} e os parâmetros populacionais com $\boldsymbol{\mu} = \mathbf{0}$, $\sigma_{11} = \sigma_{22} = ... = \sigma_{KK} = 1$ e σ_{kl}, $k \neq l$ estimados no passo M-1. Utilizando a suposição de independência entre os itens, precisaremos apenas das derivadas segundas da log-verossimilhança com relação a $\boldsymbol{\beta}_{ik}$, $i = 1, ..., n_k$, $k = 1, ..., K$. De (5.1) temos que

$$\begin{aligned}
\mathbf{H}\left(\boldsymbol{\beta}_{ik}, \boldsymbol{\beta}_{ik}\right) &= \int_{I\!R^K} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}_{ik}} \left[\frac{1}{P_{ik} Q_{ik}} \left(r_{ik} - P_{ik} f_{ik}\right) \left(\frac{\partial P_{ik}}{\partial \boldsymbol{\beta}_{ik}} \right) \right]' \right\} d\boldsymbol{\theta} \\
&= \int_{I\!R^K} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}_{ik}} \left[\frac{1}{P_{ik} Q_{ik}} \left(r_{ik} - P_{ik} f_{ik}\right) \right] \mathbf{h}'_{ik} + \frac{1}{P_{ik} Q_{ik}} \left(r_{ik} - P_{ik} f_{ik}\right) \mathbf{H}_{iik} \right\} d\boldsymbol{\theta} \\
&= \int_{I\!R^K} \left\{ -\frac{1}{\left(P_{ik} Q_{ik}\right)^2} \left(1 - 2P_{ik}\right) \mathbf{h}_{ik} \left(r_{ik} - P_{ik} f_{ik}\right) \mathbf{h}'_{ik} - \frac{1}{P_{ik} Q_{ik}} \mathbf{h}_{ik} f_{ik} \mathbf{h}'_{ik} \right. \\
&\quad \left. + \frac{1}{P_{ik} Q_{ik}} \left(r_{ik} - P_{ik} f_{ik}\right) \mathbf{H}_{iik} \right\} d\boldsymbol{\theta} \\
&= \int_{I\!R^K} \left\{ \frac{\left(r_{ik} - P_{ik} f_{ik}\right)}{\left(P_{ik} Q_{ik}\right)} \left[\mathbf{H}_{iik} - \frac{1}{\left(P_{ik} Q_{ik}\right)} \left(1 - 2P_{ik}\right) \mathbf{h}_{ik} \mathbf{h}'_{ik} \right] \right. \\
&\quad \left. - \frac{1}{P_{ik} Q_{ik}} f_{ik} \mathbf{h}_{ik} \mathbf{h}'_{ik} \right\} d\boldsymbol{\theta}. \qquad \text{(A.18)}
\end{aligned}$$

A.4 Habilidades (Moda da Posteriori de $\boldsymbol{\theta}_j$)

A derivada segunda de $\log g_j^*(\boldsymbol{\theta}_j)$, necessária para aplicação do processo iterativo, é dada por

$$\frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} = \frac{\partial^2 \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j, \boldsymbol{\beta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} + \frac{\partial^2 \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j}. \qquad \text{(A.19)}$$

De (6.2), segue que

$$\begin{aligned}
\frac{\partial^2 \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j,\boldsymbol{\beta})}{\partial\theta_{jk}^2} &= \frac{\partial}{\partial\theta_{jk}}\left(\alpha_{ik}V_{jik}P_{jik}^*Q_{jik}^*\right)\\
&= \alpha_{ik}\left[\left(\frac{\partial V_{jik}}{\partial\theta_{jk}}\right)P_{jik}^*Q_{jik}^* + \left(\frac{\partial P_{jik}^*Q_{jik}^*}{\partial\theta_{jk}}\right)V_{jik}\right].
\end{aligned}$$

Contudo,

$$\begin{aligned}
\frac{\partial V_{jik}}{\partial\theta_{jk}} &= -V_{jik}^2\left(\frac{\partial P_{jik}}{\partial\theta_{jk}}\right),\\
\frac{\partial P_{jik}^*Q_{jik}^*}{\partial\theta_{jk}} &= \left(1-2P_{jik}^*\right)\left(\frac{\partial P_{jik}}{\partial\theta_{jk}}\right).
\end{aligned}$$

Portanto,

$$\begin{aligned}
\frac{\partial^2 \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j,\boldsymbol{\beta})}{\partial\theta_{jk}^2} &= \alpha_{ik}\left[-V_{jik}^2\left(\frac{\partial P_{jik}}{\partial\theta_{jk}}\right)P_{jik}^*Q_{jik}^* + \left(1-2P_{jik}^*\right)\left(\frac{\partial P_{jik}}{\partial\theta_{jk}}\right)V_{jik}\right]\\
&= \alpha_{ik}\left[-V_{jik}^2P_{jik}^*Q_{jik}^* + \left(1-2P_{jik}^*\right)V_{jik}\right]\left(\frac{\partial P_{jik}}{\partial\theta_{jk}}\right)\\
&= \alpha_{ik}^2V_{jik}\left[-V_{jik}P_{jik}^*Q_{jik}^* + \left(1-2P_{jik}^*\right)\right]P_{jik}^*Q_{jik}^*. \qquad \text{(A.20)}
\end{aligned}$$

Temos ainda que

$$\frac{\partial^2 \log P(\mathbf{U}_{j..}|\boldsymbol{\theta}_j,\boldsymbol{\beta})}{\partial\theta_{jk}\partial\theta_{jl}} = 0, \quad k,l=1,...,K,\ k\neq l.$$

A partir de (6.3), temos

$$\frac{\partial^2 \log g(\boldsymbol{\theta}_j|\boldsymbol{\eta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_j'} = -\frac{\partial}{\partial\boldsymbol{\theta}_j}\left[\Sigma^{-1}\left(\boldsymbol{\theta}_j-\boldsymbol{\mu}\right)\right]' = -\Sigma^{-1}. \qquad \text{(A.21)}$$

Então, com as expressões (A.20) e (A.21) chegamos a (A.19). Se

$$\mathbf{f}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j\right) = \frac{\partial \log g_j^*(\boldsymbol{\theta}_j)}{\partial\boldsymbol{\theta}_j}, \quad \mathbf{H}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_j\right) = \frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_j'}$$

e $\hat{\boldsymbol{\theta}}_j^{(t+1)}$ é uma estimativa de $\boldsymbol{\theta}_j$ na iteração t, então teremos na iteração $t+1$

$$\hat{\boldsymbol{\theta}}_j^{(t+1)} = \hat{\boldsymbol{\theta}}_j^{(t)} - \left[\mathbf{H}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_j^{(t)}\right)\right]^{-1} \mathbf{f}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_j^{(t)}\right).$$

Bibliografia

[1] 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.

[2] Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações.* 14^oSINAPE - Caxambu - 24 a 28/07/00.

[3] Azevedo, C. L. N. (2000). *Métodos Estatísticos para Avaliação Educacional: Teoria Clássica de Medidas e Teoria da Resposta ao Item.* UFC e Programa Institucional de Bolsas de Iniciação Científica - PIBIC/CNPq - Relatório Final.

[4] Baker, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques.* New York: Marcel Dekker.

[5] 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.

[6] Bock, R. D. & Lieberman, M. (1970). *Fitting a response model for n dichotomously scored itens.* Psychometrika, 35, 179-197.

[7] Bock, R. D. & Aitkin, M. (1981). *Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm.* Psychometrika, 46, 433-459.

[8] Bock, R. D. & Zimowski, M. F. (1997). *Multiple Group IRT.* In *Handbook of Modern Item Response Theory.* W. J. van der Linden & R. K. Hambleton Eds. New York: Springer-Verlag.

[9] Chow, Y. S. & Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.

[10] Cordeiro, G. M. (1999). *Introdução à Teoria Assintótica*. 22^{o}Colóquio Brasileiro de Matemática - IMPA.

[11] Dempster, A. P. , Laird, N. M. & Rubin, D. B. (1977). *Maximum likelihood from incomplete data via the EM algorithm (with discussion)*. Journal of the Royal Statistical Society, Series B, 39, 1-38.

[12] DeRoos, Y. & 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.

[13] Doornik, J. A. (1999). *Object-Oriented Matrix Programming Using Ox, 3r ed.* London: Timberlake Consultants Press e Oxford: www.nuff.ox.ac.uk/Users/Doornik.

[14] Granger, C. V., Deutsch, A. & Linn, R. T. (1998). *Rash analysis of the functional independence measure (FIMTM) mastery test*. Arch. Phys. Med. Rehabil., 79, 52-57.

[15] Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications.

[16] Lord, F. M. (1980). *Applications of Item Response Theory to Pratical Testing Problems*. Hillsdale: Lawrence Erlbaum Associates, Inc.

[17] Marcoulides, G. A. Ed. (1998). *Modern Methods for Business Research*. Mahwah, NJ: Lawrence Erlbaum Reading, MA:: Addison-Wesley.

[18] Mardia, K. V., Kent J. T. & Bibby J. M. (1979). *Multivariate Analysis*. London: Academic Press.

[19] 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.

[20] Ministério da Educação e do Desporto (1998). *Sistema Nacional de Avaliação da Educação Básica: SAEB 97 - relatório técnico*. Rio de Janeiro: Fundação Cesgranrio.

[21] Mislevy, R. J. (1986). *Bayes Modal estimation in item response theory.* Psychometrika, 51, 177-195.

[22] Mislevy, R. J. (1991). *Randomization-based inference about latent variables from complex samples.* Psychometrika, **56-2**, 177-196.

[23] Neyman, J. & Scott, E. L. (1948). *Consistent estimates based on partially consistent observations.* Econometrika, 16(1), 1-32.

[24] Nojosa, R. T. (2001). *Modelos Multidimensionais na Teoria da Resposta ao Item.* Dissertação de Mestrado. Universidade Federal de Pernambuco - UFPE - Departamento de Estatística.

[25] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* New York: Wiley & Sons.

[26] Richardson, M. W. (1936). *The relationship between difficulty and the differential validity of a test.* Psychometrika, 1, 33-49.

[27] Seber, G. A. F. & Wild, C. J. (1989). *Nonlinear Regression.* New York: Wiley & Sons.

[28] Singer, J. M. & Andrade, D. F. (2000). *Analysis of Longitudinal Data.* In Handbook of Statistics, Vol. 18. P.K. Sen & C.R. Rao, eds.

[29] Stroud, A. H. & Secrest, D. (1966). *Gaussian Quadrature Fomulas.* Englewood Cliffs: Prentice-Hall.

[30] Swaminathan, H. & Gifford J. A. (1985). *Bayesian estimation in the two-parameter logistic model.* Psychometrika **50**, 349-364.

[31] Swaminathan, H. & Gifford J. A. (1986). *Bayesian estimation in the three-parameter logistic model.* Psychometrika **51**, 589-601.

[32] Tavares, H. R. (2001). *Teoria da Resposta ao Item para Dados Longitudinais.* Tese de Doutorado. São Paulo: IME/USP.

[33] Tennant, A., Hillman, M., Fear, J., Pickering, A. & Chamberlin, M. A. (1996). *Are we making the most of stanford health assessment questionnaire?* Brit. J. Rheum., 35, 574-578.

[34] Thomas, N. & Gan, N. (1997). *Generating Multiple Imputations for Matrix Sampling Data Analyzed with Item Response Models.* Journal of Educational and Behavioral Statistics, **22-4**, 425-445.

[35] Tsutakawa, R. K. & Lin, H. Y. (1986). *Bayesian estimation of item response curves.* Psychometrika **51**, 251-267.

[36] Tucker, L. R. (1946). *Maximum validity of a test with equivalent items.* Psychometrika, 11, 1-13.

[37] Valle, R. C. (1999). *Teoria da Resposta ao Item.* Dissertação de Mestrado. São Paulo: IME/USP.

[38] Van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory.* New York: Springer-Verlag

[39] Vianna, H. M. (1987). *Testes em Educação.* São Paulo: IBRASA.