

**TEORIA DA RESPOSTA AO ITEM
PARA DADOS LONGITUDINAIS**

Héilton Ribeiro Tavares

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

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

Durante a elaboração deste trabalho o autor recebeu apoio financeiro da CAPES

-São Paulo, Fevereiro de 2001-

TEORIA DA RESPOSTA AO ITEM PARA DADOS LONGITUDINAIS

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Héilton Ribeiro Tavares e aprovada pela comissão julgadora.

Comissão Julgadora:

- Prof. Dr. Dalton Francisco de Andrade (orientador) - UFC
- Prof. Dr. Julio da Motta Singer - IME/USP
- Prof. Dr. Gauss Moutinho Cordeiro - UFBA
- Prof. Dr. Clarice Garcia Borges Demétrio - ESALQ/USP
- Prof. Dr. Ruben Klein - LNCC/CNPq

*À minha esposa
Regina,
e ao meu filho
Henrique*

Agradecimentos

Eu gostaria de agradecer a todos os que, direta ou indiretamente, contribuíram para a elaboração deste trabalho. Em especial, eu gostaria de agradecer:

- Ao Prof. Dr. Dalton Francisco de Andrade, orientador e amigo, por sua dedicação e imensa paciência. Pela sua sempre presente disposição em atender e tirar dúvidas que me tiravam o sono. Pela sua mão estendida nas horas difíceis e por tudo mais.
- À minha família pela força, carinho e companheirismo sempre presentes. Primeiramente, à minha esposa Regina e ao meu filho Henrique, que em muitos momentos não puderam contar com a presença do marido e pai. À minha mãe Efigênia, meu pai Licínio, minhas irmãs Rosângela, Edivana, Silvana, Rosana e Gilvana e meus barulhentos sobrinhos Tainah, Brenno, Brunno, Érica, Felipe, Léo, Jad e Letícia.
- Ao Departamento de Estatística da Universidade Federal do Pará pela liberação para o programa, e aos amigos que lá deixei.
- Ao Departamento de Estatística do IME-USP pelo excelente programa proporcionado e aos amigos que dele fazem parte, dentre os quais detaco o Julio, Nelson, Antônio, Beti, Lúcia, Rinaldo, Clélia e Fábio.
- À Secretaria de Estado da Educação de São Paulo, em particular à Profa. Rose Neubauer, pela cessão dos dados do projeto *Avaliação do processo de inovações no ciclo básico e seu impacto sobre a situação de ensino-aprendizagem na região metropolitana de São Paulo* (AVJU).
- À CAPES pelo financiamento durante todo o programa de doutorado.
- Aos amigos Gilberto Matos e Ronald Targino pelo suporte computacional, parte fundamental na fase final do projeto.
- Aos Amigos Jairo, Sandra, Domingo, Gladys, Adrilayne, Hernan, Edwin, Cláudia, Sebastião, Celso, Júlio, Deborah, Delhi, Kátia, Laura, Chiquinho, Audrey e aos vários outros pelos papos descontraídos na hora do cafezinho e muitas bebedeiras proporcionadas.
- Aos professores Clarice Demétrio, Gauss Cordeiro, Julio Singer e Ruben Klein, membros da banca, por suas sugestões e críticas que muito contribuíram para a versão final deste trabalho.

Resumo

Neste trabalho consideramos a situação em que um grupo de indivíduos é acompanhado ao longo de várias condições de avaliação pré-fixadas. Seu desempenho (habilidade/proficiência) é obtido a partir de respostas dadas em testes formados por *itens* (questões) com características próprias. Para modelar a dependência entre suas habilidades nos vários períodos estudados são propostas diversas estruturas de covariância. São apresentadas equações de estimação para o conjunto de características dos itens e para os parâmetros da distribuição latente. É abordado o caso de dados incompletos, o caso em que há várias populações em estudo, e são apresentadas propostas para modelar as curvas de crescimento relativas aos parâmetros da distribuição latente. São feitas simulações e uma aplicação aos dados do projeto de Avaliação da Jornada Única, realizado pela Secretaria de Estado da Educação de São Paulo no período 1992-94.

ABSTRACT

In this work we consider the situation where a group of individuals is followed along two or more pre-stated conditions of evaluation. Their abilities/proficiencies are estimated from their answers to a set of tests presented to them at the different occasions. Several covariance structures are proposed to model the possible dependency between one's abilities along time. The estimation equations are presented for the items parameters and also for the abilities parameters. The incomplete data and the multiple group cases are also considered, as well as growth curve models to explain the behaviour of the abilities' location parameters along time. Simulation studies and an application to real data are presented.

Conteúdo

Agradecimentos	iv
Resumo	v
Lista de Tabelas	ix
1 Introdução	1
1.1 Conceitos Iniciais	1
1.1.1 Modelos para itens dicotômicos ou dicotomizados	2
1.1.2 Modelos para itens não-dicotômicos	4
1.2 Estimação	5
1.3 Definição da Escala de Medida	7
1.4 Modelo para Várias Populações	8
1.5 Objetivos	9
2 Modelo Longitudinal para Uma Única População: Estimação dos Parâmetros dos Itens	10
2.1 Modelo Probabilístico	10
2.2 Distribuição das Habilidades	12
2.3 Construção da Verossimilhança	12
2.4 Equações de Estimação: Caso Geral	14
2.4.1 Aplicação ao Modelo Logístico de 3 Parâmetros (ML3)	16
2.5 Implementação	18
2.5.1 Usando Suposições Adicionais	20
3 Modelo Longitudinal para Uma Única População: Estimação dos Parâmetros Populacionais	21
3.1 Equação de Estimação: Caso Geral	21
3.2 Habilidades Normalmente Distribuídas	22
3.3 Matriz de Covariância Diagonal	24
3.4 Matriz de Covariância Uniforme	27
3.5 Matriz de Covariância de Bandas	28
3.6 Matriz de Covariância AR(1)	30

3.7	Matriz de Covariância de Hankel	31
3.8	Comentários Adicionais	32
4	Modelo Longitudinal para Uma Única População: Estimação Conjunta - Parâmetros dos Itens e Populacionais	34
4.1	Implementação	35
4.1.1	Aplicação do Algoritmo EM	36
5	Modelo Longitudinal para Uma Única População: Estimação Bayesiana - Parâmetros dos Itens e Populacionais e Habilidades	39
5.1	Estimação dos Parâmetros dos Itens e Populacionais	39
5.2	Estimação das Habilidades	44
5.2.1	Moda da Distribuição Condicional de θ_j	45
5.2.2	Média da Distribuição Condicional de θ_j	47
6	Modelo Longitudinal para Uma Única População: Dados Incompletos	49
6.1	Estimação dos Parâmetros dos Itens	50
6.2	Estimação dos Parâmetros Populacionais	52
7	Modelo Longitudinal para Várias Populações	53
7.1	Modelo Probabilístico	53
7.2	Distribuição das Habilidades	54
7.3	Construção da Verossimilhança	54
7.4	Estimação dos Parâmetros dos Itens	55
7.4.1	Aplicação ao Modelo Logístico de 3 Parâmetros	56
7.5	Estimação dos Parâmetros Populacionais	57
8	Curvas de Crescimento	60
8.1	Equações de Estimação	62
8.2	Curva de Crescimento de Heitjan	64
8.3	Curva de Crescimento Polinomial	66
9	Aplicações	68
9.1	Estimação dos Parâmetros Populacionais	69
9.2	Estimação dos Parâmetros dos Itens	71
9.3	Estimação Conjunta	73
9.4	Aplicação a Dados Reais: Projeto AVJU	76
10	Comentários e Propostas para Pesquisas Futuras	79
A	Desenvolvimentos Complementares:	
	Matriz Hessiana	84

A.1 Parâmetros dos itens	84
A.2 Parâmetros populacionais	88
A.3 Parâmetros dos Itens e Parâmetros Populacionais	93
A.3.1 Expressões para o Algoritmo EM	95
A.4 Habilidades	95
A.5 Parâmetros dos Itens e Parâmetros Populacionais: Várias Populações	96
A.5.1 Expressões para o Algoritmo EM	98
A.6 Curvas de Crescimento: Caso Geral	98
A.7 Curvas de Crescimento: Heitjan	100
Referências Bibliográficas	104

Lista de Tabelas

8.1	Modelos de curvas de crescimento para o vetor de médias da população k	61
9.1	Valores dos parâmetros populacionais usados nas simulações	68
9.2	Valores adotados para os parâmetros dos itens	70
9.3	Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,25$	71
9.4	Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,75$	71
9.5	Soma dos quadrados dos desvios com relação aos verdadeiros valores. . . .	72
9.6	Soma das variâncias associadas a cada parâmetro	72
9.7	Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,25$	74
9.8	Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,75$	74
9.9	Soma dos quadrados dos desvios com relação aos verdadeiros valores. . . .	75
9.10	Soma das variâncias associadas a cada parâmetro	75
9.11	Estimativas dos Parâmetros Populacionais (erro-padrão)	77
9.12	Log-verossimilhança associada a cada estrutura de covariância	78
9.13	Soma das variâncias associadas a cada parâmetro	78

Notações e Abreviações Utilizadas no Texto

Expressões / Abreviações

t.q.: tal que
 i.e.: isto é (*id est*)
 e.g.: por exemplo (*exempli gratia*)
 v.a. : variável aleatória
 i.i.d. : independentes e identicamente distribuídas
fdp: função densidade de probabilidade
 FRI: Função de Resposta do Item
 CCI: Curva Característica do Item
 FDC: Funções Distribuições (acumuladas) de v.a. Contínuas
 EMV: Estimador de Máxima Verossimilhança
 ML*k*: Modelo Logístico de *k* parâmetros (ML1, ML2 e ML3)
 Grupo: conjunto de indivíduos da amostra

Nomenclatura

$N(0, 1)$: distribuição Normal padrão
 $N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: distribuição Normal Multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$

Parâmetros e Quantidades

N : número de indivíduos envolvidos na análise
 n : número de itens envolvidos na análise
 θ : habilidade ou proficiência
 ζ_i : parâmetros associados ao item i
 a_i : parâmetro de discriminação
 b_i : parâmetro de dificuldade
 c_i : parâmetro de acerto ao acaso
 $\boldsymbol{\zeta} = (\zeta'_1, \dots, \zeta'_n)'$: vetor integral de parâmetros
 $\boldsymbol{\eta}$: parâmetros populacionais

Símbolos

\simeq : aproximadamente
 \sim : tem distribuição
 \propto : proporcional
 \equiv : equivalente

Introdução

1.1 Conceitos Iniciais

Em muitos estudos, como potencial de compra de um certo produto em *marketing*, proficiência em determinado conteúdo em *avaliação educacional* e predisposição a uma determinada anomalia em *genética*, algumas quantidades de interesse não podem ser medidas diretamente. Variáveis desse tipo são, de forma geral, denominadas de *variáveis latentes*. Em avaliação educacional elas são referidas por *habilidade* ou *proficiência* em uma particular área de interesse, tal como ciências, matemática, trigonometria, capacidade de leitura etc. Tais quantidades são obtidas com base nos resultados de uma seqüência de ensaios, como testes formados por vários *itens* (questões). As estimações destas habilidades têm sido um dos principais objetivos dos avaliadores, e muitos trabalhos têm sido apresentados com esta finalidade.

Os primeiros esforços no sentido de introduzir uma modelagem estatística para estimação das habilidades tiveram por base os escores individuais (total de pontos em testes) e foram feitos por Spearman (1904), com forma axiomática final devida a Novick (1966). Considerando a situação em que temos N indivíduos submetidos a G testes, a proposta é decompor o escore observado do indivíduo j no teste g , T_{jg} , em duas parcelas

$$T_{jg} = \theta_j + e_{jg}, \quad j = 1, \dots, N; \quad g = 1, \dots, G,$$

onde θ_j é a habilidade e e_{jg} é um erro de medida associado ao indivíduo j . Este é, basicamente, o modelo linear de Análise de Variância com um fator fixo com N níveis e G observações para cada tratamento. Portanto, a estimação das habilidades pode ser feita via ANOVA convencional. Estas propostas iniciais foram denominadas de *Análise Clássica de Testes*. Infelizmente, todas as propostas iniciais para estimação das habilidades eram baseadas apenas nos escores individuais e não em suas partículas, ou seja, nos itens. Em 1968, Lord & Novick fundamentaram a *Análise Clássica de Itens* com o objetivo de analisar testes a partir de seus itens. Contudo, esta análise não usava muito além do que algumas estatísticas descritivas e quase nenhuma modelagem ou estatística inferencial. Além disso, a maioria das medidas usadas na análise de itens eram baseadas nos escores dos respondentes; por exemplo, medidas de correlação entre as respostas a um particular item e os escores individuais. Portanto, as conclusões para um particular item dependiam do teste

como um todo, e não somente do item considerado, logo não podiam ser estendidas para qualquer outro teste.

Uma das críticas mais fortes com relação à Teoria Clássica baseia-se no fato que se dois indivíduos A e B fazem um mesmo teste com n itens, onde o indivíduo A acerta os k itens mais fáceis e o indivíduo B acerta os k itens mais difíceis, então os indivíduos A e B têm a mesma medida de habilidade na área de conhecimento coberta pelo teste, o que não parece ser muito realista. Mesmo que seja proposta uma ponderação com relação aos itens para o cálculo desta habilidade, tal ponderação seria atribuída pelo elaborador do item ou do teste, e portanto bastante subjetiva. Para detalhes sobre a teoria clássica, veja Lord & Novick (1968).

A Teoria da Resposta ao Item (TRI) surgiu como uma forma de considerar cada item particularmente, sem relevar os escores individuais. Desta forma, as conclusões não dependem propriamente do teste, mas de cada item que o compõe. Com a TRI é possível, por exemplo, fazer várias comparações que a teoria clássica não permite, tal como comparar as habilidades de indivíduos da mesma população submetidos a testes completamente diferentes, ou ainda a comparação de indivíduos de populações diferentes submetidos a testes parcialmente distintos.

A proposta da TRI é modelar a probabilidade associada a cada *categoria* de resposta (também tratadas como *opções* ou *alternativas*), tal como certo/errado ou A/B/C/D/E. De acordo com isso, podemos dizer que os itens podem ter

- respostas dicotômicas ou dicotomizadas,
- respostas com mais de duas categorias.

A seguir, faremos uma breve descrição dos principais modelos adotados na TRI.

1.1.1 Modelos para itens dicotômicos ou dicotomizados

Os modelos apresentados nesta seção podem ser utilizados tanto para a análise de itens dicotômicos ou de múltipla escolha dicotomizados (corrigidos como certo ou errado) quanto para a análise de itens abertos (de resposta livre), quando avaliados de forma dicotomizada.

Consideremos N indivíduos submetidos a um teste com n itens de múltipla escolha, dicotômicos ou dicotomizados. A resposta do indivíduo j , $j = 1, \dots, N$, ao item i , $i = 1, \dots, n$, será representada pela variável aleatória U_{ji} , com $U_{ji} = 1$ (*correta*) e $U_{ji} = 0$ (*incorreta*). A proposta da TRI é modelar a distribuição de U_{ji} , ou equivalentemente, a probabilidade de acerto ao item i , para cada item. A *habilidade* do indivíduo j passa a constar no modelo como um parâmetro e é, geralmente, representada por θ_j .

Claramente, a probabilidade de acerto ao item i deverá ser representada por uma função monótona crescente de θ_j com imagem no intervalo $[0,1]$, o que sugere a escolha dessa função na classe de funções de distribuições (acumuladas) de variáveis aleatórias contínuas

(FDC). Essas funções são, freqüentemente, denotadas por Ψ_i e denominadas *Função de Resposta do Item (FRI)* (Van der Linder & Hambleton, 1997) ou *Curva Característica do Item (CCI)* (Tucker, 1946). Muitas outras funções fora dessa classe foram propostas, tal como a função linear, mas com pouco impacto prático e sérias limitações nos valores possíveis para as habilidades. Por outro lado, situações práticas freqüentemente indicavam um ajuste insatisfatório dos modelos causado por uma assíntota ($\theta \rightarrow -\infty$) não nula, sugerindo que indivíduos com baixa habilidade respondiam corretamente ao item i com probabilidade c_i positiva. Isso foi interpretado pelos especialistas como acertos ao acaso (chutes). Por conta disso, uma outra classe de funções bem aceita pode ser escrita como $P_i = c_i + (1 - c_i)\Psi_i$.

Em 1952, Lord propôs o Modelo Normal

$$\Psi_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

onde a_i e b_i são parâmetros relativos ao item i . Em 1957, Birnbaum propôs o modelo

$$\Psi_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}},$$

onde D é um fator de escala. O modelo acima é denominado de *Modelo Logístico* quando $D = 1$. O valor 1.702 para D é utilizado quando se deseja que os valores para as funções Normais e Logísticas sejam aproximadamente iguais (ver Haley, 1952). Na prática, o modelo mais utilizado é o logístico com 3 parâmetros, descrito a seguir.

Modelo logístico de 3 parâmetros (ML3)

O modelo logístico (*unidimensional*) de 3 parâmetros é dado por:

$$P(U_{tw} = 1 | \theta_j, \zeta_i) = c_i + (1 - c_i) \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1}.$$

onde $\zeta_i = (a_i, b_i, c_i)'$ é o vetor de parâmetros associado ao item i (ou simplesmente, *parâmetros do item i*), com $a_i > 0$, $b_i \in (-\infty, \infty)$, $c_i \in [0, 1]$, D é uma constante de escala e $\theta_j \in (-\infty, \infty)$ é a habilidade do indivíduo j . Sua representação gráfica está na Figura 1.1.

De um modo geral, podemos dizer que o parâmetro b_i está relacionado à dificuldade do item, quanto maior este parâmetro mais difícil será o item, e por isso ele é conhecido como *parâmetro de dificuldade*. Aumentando o valor de b_i estaremos deslocando a FRI para a direita e, portanto, um indivíduo com habilidade fixada terá menor probabilidade de acertar o item, ou seja, o item torna-se mais difícil. O parâmetro c_i nos dá a probabilidade de um indivíduo com baixa habilidade responder corretamente ao item, e é também chamado de parâmetro de *acerto ao acaso*. Para itens com m alternativas (4 ou 5, em geral), normalmente espera-se que c_i esteja próximo de $1/m$, já que se supõe que os indivíduos com baixa habilidade respondem ao acaso esse item. Se obtivermos $c_i \gg 1/m$, há indicação

Figura 1.1 *Representação gráfica do ML3*

que algumas das m alternativas foram facilmente eliminadas, mesmo por indivíduos de baixa habilidade. Lord (1974) argumenta que o caso $c_i \ll 1/m$ é mais comum devido à presença de algum *distractor* (alternativa incorreta) que esteja atraindo indivíduos com baixa habilidade.

Podemos notar ainda, fazendo $\theta_j = b_i$, que b_i é a habilidade necessária para que a probabilidade de acerto ao item seja $(1+c_i)/2$. Ainda, o coeficiente angular da tangente da FRI em $\theta_j = b_i$ é $D(1-c_i)a_i/4$, ou seja, é proporcional a a_i . Logo, quanto maior o valor de a_i , mais íngreme será a FRI e se considerarmos dois indivíduos com habilidades $\theta_1 = b_i - \delta$ e $\theta_2 = b_i + \delta$, $\delta > 0$, então a diferença entre $P_i(\theta_2)$ e $P_i(\theta_1)$ (onde $P_i(\theta_j) = P(U_{ji} = 1 | \theta_j, \zeta_i)$) será crescente com a_i , ou seja, o aumento de a_i leva a uma maior discriminação entre os dois indivíduos, e por isso ele é denominado de *parâmetro de discriminação do item*.

Em algumas situações, pode-se supor que não há possibilidade de respostas ao acaso (ie., $c_i = 0$), resultando no *Modelo Logístico de 2 parâmetros (ML2)*. Se, além disso, pudermos supor que o poder de discriminação é o mesmo para todos os itens envolvidos em um teste, isto é, $a_i = a, \forall i$, chegamos ao *Modelo Logístico de 1 parâmetro (ML1)*, onde a constante a é incorporada na *unidade de medida* da habilidade (ver Seção 1.3).

1.1.2 Modelos para itens não-dicotômicos

Vários autores defendem a idéia de que a dicotomização elimina informações associadas a itens respondidos de forma incorreta, ou seja, sabe-se apenas que um particular indivíduo respondeu incorretamente ao item, mas não é relevada a informação sobre a alternativa escolhida. Por outro lado, alguns itens têm a característica que suas alternativas são ordenadas, e isso exige a introdução de novos modelos que incorporem essa informação. Para maiores detalhes sobre esses modelos, veja Hambleton & Swaminathan (1985) e Andrade, Tavares & Valle (2000).

Vários outros modelos de resposta têm sido propostos recentemente. Dentre eles, pode-

mos destacar os modelos multidimensionais (ver Ackerman, 1996), modelos não-paramétricos e modelos não-monotônicos (ver Van der Linden & Hambleton, 1997).

Embora exista uma vasta classe de estruturas de itens, podemos dizer que os itens dicotômicos ou dicotomizados são, sem dúvida, os de maior interesse prático. Por conta disso, o restante deste texto é dedicado somente a esse tipo de item. Além disso, trataremos apenas da classe de modelos escrita como $P_i = c_i + (1 - c_i)\Psi_i$, onde Ψ_i é uma FDC.

1.2 Estimação

Consideremos a situação em que temos N indivíduos submetidos a um teste com n itens. Nesta situação, o objetivo de uma análise pode ser estimar as habilidades individuais, estimar parâmetros relativos aos itens, ou ambos. Em algumas situações podemos pensar que os itens já foram *calibrados* (isto é, seus parâmetros já foram estimados) e desejamos simplesmente estimar as habilidades individuais. Por outro lado, indivíduos com as habilidades estimadas em um teste poderão ser utilizados na calibração de um novo conjunto de itens. Em geral, os objetivos são tanto a calibração dos itens como a estimação das habilidades.

Quando temos apenas uma população em estudo, o objetivo principal se concentra em estimar os parâmetros dos itens, ζ_i , $i = 1, \dots, n$, e as habilidades individuais, θ_j , $j = 1, \dots, N$. Geralmente, essa estimação é feita pelo Método da Máxima Verossimilhança através da aplicação de algum processo iterativo, como *Newton-Raphson* [ver Issac & Keller (1966), por exemplo] ou “*Scoring*” de Fisher [ver Rao (1973), por exemplo], com duas abordagens usuais: *estimação conjunta* (parâmetros dos itens e habilidades) ou *em duas etapas* (primeiro a estimação dos parâmetros dos itens e, posteriormente, das habilidades). No caso da estimação conjunta, o número de parâmetros pode ser muito grande ($3n + N$ para o ML3), levando a uma enorme exigência computacional que envolve a inversão de matrizes dessa ordem. Para contornar esse problema, Birnbaum (1968) propôs um processo *vai e vem* (“back-and-forth”), que é iniciado com estimativas grosseiras das habilidades (escores padronizados, por exemplo) e envolve a estimação dos parâmetros dos itens considerando as habilidades conhecidas; após a obtenção das estimativas dos parâmetros dos itens, as habilidades são estimadas considerando os parâmetros dos itens conhecidos. Esses passos são repetidos até que algum critério de parada do processo seja alcançado. A grande vantagem desse método é que os itens (i.e., seus parâmetros) são estimados individualmente, o que exige o tratamento de matrizes 3×3 para o ML3; esse procedimento é baseado em uma suposição bastante freqüente na TRI, de que um indivíduo com habilidade fixada θ responde aos itens de forma independente (denominada *Independência Local* ou *Independência Condicional*). As habilidades também são estimadas individualmente e, com isso, a exigência computacional diminui drasticamente. Entretanto, esse procedimento têm um problema sério: sabe-se que, para os parâmetros dos itens conhecidos, os estimadores de máxima verossimilhança (EMV) das habilidades,

$\hat{\theta}_j$, $j = 1, \dots, N$, convergem (ver Sen & Singer (1993), por exemplo) para os seus verdadeiros valores quando o número de itens cresce; com as habilidades conhecidas, os EMV dos parâmetros dos itens, $\hat{\zeta}_i$, convergem para os seus verdadeiros valores quando o número de indivíduos cresce. Na estimação conjunta, as habilidades são denominadas *parâmetros incidentais*, pois o número destes parâmetros (θ_j , $j = 1, \dots, N$) cresce com o número de indivíduos; os parâmetros dos itens são denominados *parâmetros estruturais* e o número desses parâmetros não se altera quando a amostra cresce. Essas denominações são devidas a Neyman & Scott (1948), que notaram, em um contexto diferente do da TRI, que na presença de parâmetros incidentais, o EMV dos parâmetros estruturais pode ser assintoticamente viesado. Esse problema de falta de consistência dos estimadores dos parâmetros dos itens (ou habilidades) na presença de um número muito grande de indivíduos (ou itens) foi notado por Anderson (1973) e demonstrado para o ML1. Porém, quando tanto o número de itens quanto o número de indivíduos crescem, os EMV dos parâmetros dos itens e das habilidades podem ser não-viciados, como sugerido por Lord (1968) e demonstrado apenas para o ML1 por Haberman (1975). Resultados numéricos obtidos por Lord (1975) e Swaminathan & Gifford (1983), para modelos diferentes do ML1, reforçam a conjectura de que os EMV dos parâmetros dos itens e das habilidades são não-viciados, quando o número de itens e o número de indivíduos crescem.

O problema de possível inconsistência dos estimadores obtidos em uma etapa levou ao desenvolvimento da estimação em duas etapas por Bock & Lieberman (1970). A primeira etapa consiste na estimação dos parâmetros dos itens e a segunda na estimação das habilidades. [Notemos que apesar de as habilidades individuais serem parâmetros, um conjunto destes valores pode ser representado por uma distribuição, e esta ser usada para fins convenientes]. A estimação dos parâmetros dos itens é feita por *Máxima Verossimilhança Marginal*, considerando que a habilidade da população (θ) segue uma distribuição cuja função densidade de probabilidade (*fdp*) é $g(\theta|\boldsymbol{\eta})$, onde $\boldsymbol{\eta}$ é o conjunto de parâmetros associados à população, e integrando a função de verossimilhança com relação a θ . Após a estimação dos parâmetros dos itens, as habilidades são estimadas individualmente por máxima verossimilhança, ou ainda pela moda ou média da distribuição condicional de θ dado $\mathbf{U}_j = (U_{j1}, \dots, U_{jn})$ com ζ_i , $i = 1, \dots, n$, conhecidos. Embora este método tenha a vantagem de envolver, na primeira etapa, apenas a estimação dos parâmetros dos itens, a estimação é feita através de aplicação de métodos numéricos que dependem das derivadas segundas da log-verossimilhança com relação a ζ_i e ζ_k , $i, k = 1, \dots, n$, que podem ser não nulas para $i \neq k$. Com isso, há a necessidade da inversão de matrizes de ordem $3n \times 3n$ para o ML3, o que ainda pode ser bastante exigente do ponto de vista computacional. Para contornar esse problema, Bock & Aitkin (1981) reformularam as equações obtidas por Bock & Lieberman e propuseram que as estimativas de máxima verossimilhança fossem obtidas via algoritmo *EM* (ver Dempster, Laird & Rubin, 1977); essa nova proposta possibilita que as derivadas segundas citadas acima para $i \neq k$ sejam nulas. Com isso, a matriz $3n \times 3n$ (ML3) de derivadas segundas torna-se bloco-diagonal, o que possibilita que os itens (i.e., seus parâmetros) sejam estimados individualmente.

1.3 Definição da Escala de Medida

A métrica em que os parâmetros dos itens e/ou habilidades são estimados é um dos pontos de primordial importância na fase de interpretação. Quando as habilidades são conhecidas e desejamos estimar apenas os parâmetros dos itens, estes são estimados na mesma escala das habilidades. Na situação oposta, o conhecimento dos parâmetros dos itens implica o conhecimento da escala em que eles foram estimados, de forma que as habilidades serão estimadas na mesma escala dos parâmetros dos itens.

No caso da estimação conjunta, não há nenhuma escala definida. Teoricamente, as habilidades podem assumir qualquer valor real, sem referência ao que é grande ou pequeno. Para contornar esse problema podemos escolher um ponto de referência μ e uma medida de dispersão σ . Em cada etapa do processo iterativo as estimativas das habilidades são padronizadas de forma que tenham média μ e desvio-padrão σ . Frequentemente adota-se $\mu = 0$ e $\sigma = 1$, e denota-se por $(0;1)$ ou $(0,1)$. No caso da estimação em duas etapas, geralmente supõe-se que $\theta \sim N(0,1)$. Em ambos os casos, estaremos fixando a *escala das habilidades* e, conseqüentemente, a dos parâmetros dos itens, no par $(0,1)$, onde o primeiro número representa a média e o segundo o desvio-padrão da distribuição da habilidade da população considerada no estudo.

Rigorosamente, a escala dos parâmetros é fixada para eliminar a *não-identificabilidade* do modelo. Essa não-identificabilidade ocorre porque mais de um conjunto de parâmetros produz o mesmo valor da FRI, e conseqüentemente, da verossimilhança. Por exemplo, para o ML2 e o ML3, se $\theta_j^* = \alpha\theta_j + \beta$, $b_i^* = \alpha b_i + \beta$, $a_i^* = a_i/\alpha$ e $c_i^* = c_i$, onde α e β são constantes reais com $\alpha > 0$, então $P(U_{ji} = 1 | \theta_j, \zeta_i) = P(U_{ji} = 1 | \theta_j^*, \zeta_i^*)$. Esta não-identificabilidade pode ser eliminada de várias formas, sendo que as restrições mais comuns são as citadas anteriormente. Para maiores detalhes, veja Baker (1992).

Quando consideramos mais de uma população (relativas a diferentes graus de escolaridade, sexo, grupos étnicos, regionais etc.) a estimação é realizada em separado para cada uma das populações. Embora cada uma destas populações (e os respectivos itens) tenham seus parâmetros estimados em uma escala $(0,1)$, estas escalas não são comparáveis porque cada uma está vinculada a uma população. Daí, surge a necessidade de colocar tais parâmetros na mesma escala, ou seja, fazer o que é denominado de *Equalização* (“Equating”). Esse processo exige que os testes aplicados às várias populações tenham itens comuns fazendo uma *ligação* entre eles. Por exemplo, se tivermos 3 populações (i.e., testes) devemos ter pelo menos itens comuns entre os testes 1 e 2 e também entre os testes 2 e 3 (ou 1 e 3). As equalizações são feitas para cada par de testes com itens comuns, mas como há uma ligação entre todos os testes, o produto final é a equalização de todos eles. Entretanto, antes de proceder com a equalização entre duas populações é necessário estabelecer qual população será adotada como *População Referência*, para que os parâmetros relativos à outra população possam ser transformados, através de alguma função f , para a escala da população referência. Os modelos Normal e Logístico têm uma certa propriedade de linearidade (de f) entre escalas: se $\zeta_i^A = (a_i^A, b_i^A, c_i^A)'$ representa

um item na análise envolvendo a população referência (digamos, A), θ_j^A a habilidade do indivíduo j dessa população e $\zeta_i^B = (a_i^B, b_i^B, c_i^B)'$, com $c_i^B = c_i^A$, o mesmo item em outra população (digamos, B), então, a habilidade do mesmo indivíduo na escala da população B , θ_j^B , será obtida como função de θ_j^A por

$$\begin{aligned} P(U_{ji} = 1 | \theta_j^B, \zeta_i^B) &= P(U_{ji} = 1 | \theta_j^A, \zeta_i^A) \\ \Leftrightarrow a_i^B(\theta_j^B - b_i^B) &= a_i^A(\theta_j^A - b_i^A) \\ \Leftrightarrow \theta_j^B &= (a_i^A/a_i^B)\theta_j^A + [b_i - (a_i^A/a_i^B)b_i^A] \\ &= \alpha\theta_j^A + \beta, \quad \text{com } \alpha > 0 \text{ e } \beta \in \mathbb{R}. \end{aligned}$$

Esta mesma relação é válida para o parâmetro de dificuldade, que também é uma habilidade: $b_i^B = \alpha b_i^A + \beta$, bem como $a_i^B = a_i^A/\alpha$. De acordo com isso e com base nas estimativas dos parâmetros de dificuldade dos itens comuns aos dois testes, obtemos uma reta de regressão entre os dois conjuntos de estimativas e, assim, chegamos à relação entre os parâmetros nas duas populações. Isso possibilita que todas as habilidades e os parâmetros dos itens associados a uma população sejam transformados para a escala da população referência. Isso pode ser feito para todas as populações consideradas com base nos itens que fizeram as ligações entre os testes.

O método de equalização descrito acima é conhecido como método da *Regressão*, mas existem vários outros métodos, dentre os quais podemos destacar: *Média-Desvio (Mean-Sigma)*, *Média-Desvio Robusto* e *Curva Característica*. Para detalhes, veja Kolen & Brennan (1995) e Andrade (1999).

1.4 Modelo para Várias Populações

O modelo para várias populações foi proposto por Bock & Zimowski (1997) e representou um grande avanço na TRI (ver Hedges & Vevea, 1997). Nesse modelo, consideram-se K populações independentes em estudo e é feita uma análise conjunta das respostas amostrais dessas populações. Considera-se que a habilidade da população k segue uma determinada distribuição cuja média é μ_k e variância σ_k^2 , $k = 1, \dots, K$. Frequentemente, adota-se a distribuição Normal, isto é, $\theta_k \sim N(\mu_k, \sigma_k^2)$, $k = 1, \dots, K$. Novamente, existe a necessidade do arbítrio de uma população referência cujas habilidades serão dadas por uma $N(0, 1)$.

A grande vantagem da abordagem de Bock & Zimowski está no fato de a equalização ser feita automaticamente no próprio processo de estimação. Desta forma, não estaremos mais sujeitos a diferenças nas estimativas dos parâmetros devidas ao método de equalização escolhido. Além disso, na presença de várias populações (digamos, $K \geq 5$), com as equalizações sendo feitas entre os testes k e $k + 1$, temos erros (relativos à regressão, por exemplo) associados a cada equalização entre duas populações, que serão acumulados

para a estimação de (μ_2, σ_2^2) , (μ_3, σ_3^2) , \dots e, principalmente, de (μ_K, σ_K^2) . Portanto, estes parâmetros poderão ser muito mal estimados.

1.5 Objetivos

Uma suposição adotada na modelagem de Bock & Zimowski é que as populações são independentes, induzindo que cada indivíduo só deve compor uma população e, portanto, tornando o uso desta teoria inapropriada em conjuntos de dados oriundos de estudos longitudinais.

Neste trabalho são propostos modelos da Teoria da Resposta ao Item para Dados Longitudinais que permitem a incorporação de estruturas de correlação em suas formulações. Consideramos que os dados possuem uma estrutura longitudinal, ou seja, os indivíduos envolvidos nas avaliações podem ser submetidos a vários testes, em uma mesma área do conhecimento, ao longo de T condições de avaliação (tempo), tais como diferentes séries escolares. Neste contexto, o modelo sugerido por Bock & Zimowski passa a ser um caso particular dos modelos aqui propostos.

O Capítulo 2 é dedicado à estimação dos parâmetros dos itens, considerando que os *parâmetros populacionais* (matriz de covariâncias e vetor de médias) são conhecidos. O Capítulo 3 é dedicado à estimação dos parâmetros populacionais, com os parâmetros dos itens conhecidos. O Capítulo 4 completa os Capítulos 2 e 3 na situação em que temos a estimação conjunta dos parâmetros dos itens e parâmetros populacionais. O Capítulo 5 trata da estimação bayesiana, tantos dos parâmetros dos itens e populacionais, como das habilidades. Nestes capítulos, 2 a 5, consideramos a existência de uma única população com a estrutura dos dados completa. No Capítulo 6 abordamos o caso em que há dados incompletos, ainda com uma única população. Em seguida, no Capítulo 7, consideramos o caso em que temos várias populações, com várias condições de avaliação, para dados completos e incompletos. No Capítulo 8, com um perfil um pouco diferente dos anteriores, tratamos de curvas de crescimentos associadas aos vetores de médias. Finalmente, no Capítulo 9 são feitas aplicações a dados simulados e a dados reais. Concluimos com algumas considerações gerais, deixando desenvolvimentos teóricos adicionais para os apêndices.

Modelo Longitudinal para Uma Única População: Estimação dos Parâmetros dos Itens

Neste capítulo construímos um modelo considerando a existência de uma única população de respondentes e T períodos de avaliação para dados completos e parâmetros populacionais conhecidos. Por facilidade de exposição, consideramos que as avaliações são realizadas nos tempos $1, 2, \dots, T$, e nos referimos a estas avaliações como *Teste 1*, *Teste 2*, \dots , *Teste T*. Desta forma, usaremos indistintamente os termos *teste*, *tempo* e ainda *período (condição) de avaliação*. O conjunto de indivíduos que compõem a amostra, obtida por amostragem aleatória simples, será referido por *grupo*. Consideramos que o teste t é formado por n_t itens, e que temos n itens no total dos T testes, com $n \leq \sum_{t=1}^T n_t$. Assumimos uma determinada ordenação no conjunto dos n itens, de forma que possamos denotá-los (i.e., seus parâmetros) por $\zeta = (\zeta'_1, \zeta'_2, \dots, \zeta'_n)'$.

A situação mais comum na prática é aquela que temos que estimar tanto os parâmetros dos itens como os populacionais. Esta estimação geralmente é dividida em dois blocos, onde o primeiro se refere à estimação dos parâmetros dos itens, considerando fixos os parâmetros populacionais, e o segundo bloco se destina à estimação dos parâmetros populacionais, considerando fixos os parâmetros dos itens. Neste capítulo tratamos somente do primeiro bloco, o próximo capítulo trata do segundo bloco e no Capítulo 4 fazemos a junção dos dois blocos. Vale notar que se estivermos na situação em que os parâmetros populacionais realmente já forem conhecidos, apenas os parâmetros dos itens precisam ser estimados e, portanto, apenas o primeiro bloco será aplicado.

2.1 Modelo Probabilístico

A princípio, qualquer um dos modelos propostos na literatura [ver Andrade, Tavares & Valle (2000), por exemplo] pode ser usado para fins de aplicação. Neste trabalho, adotaremos o *Modelo de Birnbaum* ou *Modelo Logístico de 3 parâmetros (ML3)* para a probabilidade de um indivíduo responder corretamente a um item. Este modelo tem sido o mais utilizado por profissionais das áreas de avaliação. Além disso, como descrito no Capítulo 1, com ele é possível obter alguns sub-modelos (ML2 e ML1) e ainda, estimativas dos parâmetros dos itens muito próximas daquelas obtidas com o Modelo Normal.

Se U_{jit} é a resposta (binária) ao item i oriunda do j -ésimo indivíduo no t -ésimo teste, então

$$P(U_{jit} = 1 | \theta_{jt}, \zeta_i) = c_i + (1 - c_i) \{1 + e^{-Da_i(\theta_{jt} - b_i)}\}^{-1}, \quad (2.1)$$

onde θ_{jt} é a habilidade do indivíduo j no teste t e $\zeta_i = (a_i, b_i, c_i)'$ é o vetor de parâmetros associado ao item i .

Vamos assumir que a probabilidade de acerto a qualquer item, em qualquer dos testes, pode ser modelada pela mesma função de resposta (ML3). Esta suposição é freqüentemente aceitável na prática e facilita bastante o desenvolvimento da teoria.

Uma suposição que tem sido freqüente na literatura, e que será usada neste trabalho, é a de *Independência Local* (ou *Independência Condicional*), na qual os itens que compõem cada teste são respondidos de forma independente por cada indivíduo com habilidade θ , ou seja,

$$P(U_{j1t} = u_{j1t}, \dots, U_{jn_t t} = u_{jn_t t} | \theta_t, \zeta) = \prod_{i=1}^{n_t} P(U_{jit} = u_{jit} | \theta_t, \zeta_i), \quad t = 1, \dots, T.$$

Em outras palavras, a habilidade de um indivíduo é informação suficiente para determinar a probabilidade de acerto de um determinado item (conhecidos seus parâmetros), não importando se ele respondeu corretamente ou não a outros itens. Cabe ressaltar que não estamos supondo que as respostas aos itens são independentes, mas sua dependência se dá apenas através da habilidade. Por conta disso, fixada a habilidade de um indivíduo, suas respostas aos diferentes itens serão condicionalmente independentes.

Estenderemos a propriedade de independência local ao que chamaremos de *Independência Local-Temporal* (*Local – Longitudinal*), que declara que os itens de um teste serão respondidos independentemente dos itens dos outros testes, fixadas as habilidades nos T testes. Mais ainda, assumiremos que as respostas de cada indivíduo aos itens do teste t dependem apenas da habilidade neste teste, θ_{jt} , e não das habilidades nos outros $T - 1$ testes, ou seja,

$$P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \zeta) = \prod_{t=1}^T P(\mathbf{U}_{j.t} | \boldsymbol{\theta}, \zeta),$$

onde $\mathbf{U}_{j..}$ representa o vetor de respostas do indivíduo j em todos os tempos, enquanto $\mathbf{U}_{j.t}$ representa apenas as respostas do tempo t .

Também estaremos assumindo que indivíduos diferentes responderão aos itens de forma independente.

Várias outras suposições costumam ser feitas. As principais são: (i) a de que o *tempo destinado ao teste* é suficiente para que todos os indivíduos 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, referido por *unidimensionalidade*.

2.2 Distribuição das Habilidades

Assumiremos que o vetor de habilidades da população nas T condições de avaliação, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_T)'$, tem distribuição contínua multivariada com vetor de parâmetros $\boldsymbol{\eta}$ de componentes finitas, e denotaremos sua *fdp* por $g(\boldsymbol{\theta}|\boldsymbol{\eta})$, onde $\boldsymbol{\eta}$ é conhecido.

No caso Normal T -variado temos que $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, onde $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_T)'$ é o vetor de médias e $\boldsymbol{\Sigma} = (\sigma_{ts})_{t,s \leq T}$ é a matriz de covariâncias. Para os elementos da diagonal de $\boldsymbol{\Sigma}$, usaremos a notação $\sigma_{tt} = \sigma_t^2$.

Basicamente, o processo de estimação dos parâmetros dos itens é o de máxima verossimilhança. Portanto, o nosso primeiro passo será a construção da *função de verossimilhança*, ou simplesmente *verossimilhança*.

2.3 Construção da Verossimilhança

Consideremos a situação em que temos n_t itens no teste t e n itens no conjunto de todos os testes. Tal como no modelo de Bock & Zimowski, há necessidade de termos itens comuns criando uma ligação entre os testes, e por isso $n \leq n_c = \sum_{t=1}^T n_t$. Sejam

$\mathbf{U}_{j.t} = (U_{j1t}, U_{j2t}, \dots, U_{jn_t t})'$ o vetor ($n_t \times 1$) de respostas do indivíduo j no teste t ;

$\mathbf{U}_{j..} = (\mathbf{U}'_{j.1}, \mathbf{U}'_{j.2}, \dots, \mathbf{U}'_{j.T})'$ o vetor ($n_c \times 1$) de respostas do indivíduo j em todos os testes.

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

\mathbf{I}_t o conjunto dos índices dos itens presentes no teste t .

τ_i o conjunto dos índices dos testes que contém o item i .

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

$$P(\mathbf{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) = \prod_{i \in \mathbf{I}_t} P(U_{jit}|\theta_t, \zeta_i), \quad (2.2)$$

Adicionalmente, com a independência temporal podemos escrever

$$P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) = \prod_{t=1}^T P(\mathbf{U}_{j.t}|\boldsymbol{\theta}, \boldsymbol{\zeta}) = \prod_{t=1}^T P(\mathbf{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) \quad (2.3)$$

$$= \prod_{t=1}^T \prod_{i \in \mathbf{I}_t} P(U_{jit}|\theta_t, \zeta_i). \quad (2.4)$$

A probabilidade marginal do vetor de respostas $\mathbf{U}_{j..}$ é dada por

$$\begin{aligned} P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) &= \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}. \end{aligned} \quad (2.5)$$

Nesta expressão usamos que a distribuição de $\mathbf{U}_{j..}$ é função apenas de $\boldsymbol{\theta}$ e $\boldsymbol{\zeta}$, não dependendo de $\boldsymbol{\eta}$. Com (2.5) podemos escrever a verossimilhança $L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = P(\mathbf{U}_{...}; \boldsymbol{\zeta}, \boldsymbol{\eta})$ como

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \prod_{j=1}^N P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}). \quad (2.6)$$

Esta é a função que pode ser adotada para obtenção das estimativas de máxima verossimilhança. Entretanto, na literatura tem sido freqüente a construção da verossimilhança usando a abordagem de *padrão de resposta*. Como temos n_t itens no teste t , totalizando n_c respostas para cada indivíduo, com 2 possíveis respostas para cada item (0 ou 1), há um total de $S = 2^{n_c}$ possíveis respostas (*padrões de resposta*). Quando o número de indivíduos é grande com relação ao número de itens, pode haver vantagens computacionais em trabalhar com o número de ocorrências dos diferentes padrões de resposta. Neste sentido, daqui em diante vamos trabalhar considerando os padrões de resposta e não mais os vetores de respostas individuais, ou seja, índice j não mais representará um indivíduo, mas sim um padrão de resposta.

Sejam r_j o número de ocorrências do padrão de resposta j , e ainda $s \leq \min(N, S)$ o número de padrões de resposta com $r_j > 0$. Trabalhando com s estaremos excluindo todos os padrões de resposta com ocorrência nula, o que computacionalmente pode ser bastante vantajoso. Vale notar que

$$\sum_{j=1}^s r_j = N, \quad (2.7)$$

e para utilizar os vetores de respostas individuais ao invés dos padrões de resposta, basta fazermos $s = N$ e $r_j = 1, \forall j$.

Seja $\mathbf{R} = (r_1, r_2, \dots, r_s)'$ o vetor de freqüências dos padrões de resposta. Usando a independência entre as respostas dos diferentes indivíduos, teremos que \mathbf{R} segue uma distribuição *Multinomial*, isto é,

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

Desta forma, temos que a log-verossimilhança será

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

2.4 Equações de Estimação: Caso Geral

As equações de estimação para os parâmetros dos itens são

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} = \mathbf{0}; \quad i = 1, 2, \dots, n. \quad (2.8)$$

Para especificarmos essas equações, notemos que

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left\{ \sum_{j=1}^s r_j \log P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta}) \right\} \\ &= \sum_{j=1}^s r_j \frac{1}{P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})} \frac{\partial P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i}. \end{aligned} \quad (2.9)$$

Mas

$$\frac{\partial P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} = \frac{\partial}{\partial \zeta_i} \left\{ \int_{\mathbb{R}^T} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta} \right\} \quad (2.10)$$

$$= \int_{\mathbb{R}^T} \left\{ \frac{\partial}{\partial \zeta_i} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) \right\} g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta} \quad (2.11)$$

$$= \int_{\mathbb{R}^T} \left\{ \frac{\partial}{\partial \zeta_i} \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) \right\} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}$$

$$= \int_{\mathbb{R}^T} \left\{ \frac{\partial}{\partial \zeta_i} \sum_{t=1}^T \sum_{l \in \mathbf{I}_t} \log P(U_{jlt} | \theta_t, \zeta_l) \right\} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}$$

$$= \int_{\mathbb{R}^T} \left\{ \sum_{t=1}^T \sum_{l \in \mathbf{I}_t} \frac{\partial}{\partial \zeta_i} \log P(U_{jlt} | \theta_t, \zeta_l) \right\} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}$$

$$= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} \left(\frac{\partial P(U_{jlt} | \theta_t, \zeta_i) / \partial \zeta_i}{P(U_{jlt} | \theta_t, \zeta_i)} \right) \right\} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}, \quad (2.12)$$

onde a ordem da derivada e da integral em (2.10) pôde ser permutada com base no

Teorema da Convergência Dominada de Lebesgue (Chow & Teicher, 1978). Mislevy (1984) argumenta que essa permuta é amplamente satisfeita por modelos de teoria de resposta ao item e funções densidade.

Escrevendo

$$P(U_{jit}|\theta_t, \zeta) = P(U_{jit} = 1|\theta_t, \zeta)^{U_{jit}} P(U_{jit} = 0|\theta_t, \zeta)^{1-U_{jit}} = P_{it}^{U_{jit}} Q_{it}^{1-U_{jit}},$$

teremos

$$\begin{aligned} \frac{\partial}{\partial \zeta_i} P(U_{jit}|\theta_t, \zeta_i) &= \frac{\partial}{\partial \zeta_i} \left(P_{it}^{U_{jit}} Q_{it}^{1-U_{jit}} \right) \\ &= U_{jit} P_{it}^{U_{jit}-1} \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) Q_{it}^{1-U_{jit}} + (1 - U_{jit}) Q_{it}^{-U_{jit}} \left(-\frac{\partial}{\partial \zeta_i} P_{it} \right) P_{it}^{U_{jit}} \\ &= \left[U_{jit} P_{it}^{U_{jit}-1} Q_{it}^{1-U_{jit}} - (1 - U_{jit}) Q_{it}^{-U_{jit}} P_{it}^{U_{jit}} \right] \left(\frac{\partial P_{it}}{\partial \zeta_i} \right). \end{aligned}$$

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

$$\frac{\partial}{\partial \zeta_i} P(U_{jit}|\theta_t, \zeta_i) = (-1)^{U_{jit}+1} \left(\frac{\partial P_{it}}{\partial \zeta_i} \right). \quad (2.13)$$

Notando agora que

$$\frac{(-1)^{U_{jit}+1} P_{it} Q_{it}}{P_{it}^{U_{jit}} Q_{it}^{1-U_{jit}}} = \begin{cases} Q_{it} & \text{se } U_{jit} = 1 \\ -P_{it} & \text{se } U_{jit} = 0, \end{cases} \quad (2.14)$$

podemos reescrever este termo como $U_{jit} - P_{it}$. Segue de (2.12) que

$$\frac{\partial P(\mathbf{U}_{j..}|\zeta, \eta)}{\partial \zeta_i} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} \frac{(U_{jit} - P_{it})}{P_{it} Q_{it}} \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) \right\} P(\mathbf{U}_{j..}|\theta, \zeta) g(\theta|\eta) d\theta. \quad (2.15)$$

Como comentado no Capítulo 1, vamos nos restringir à classe de FRI escritas como $P_{it} = c_i + (1 - c_i)P_{it}^*$, onde P_{it}^* representa uma FDC. Para os casos Logístico e Normal, temos

$$P_{it}^* = \begin{cases} \{1 + e^{-Da_i(\theta_t - b_i)}\}^{-1}, & \text{Modelo Logístico,} \\ \int_{-\infty}^{a_i(\theta_t - b_i)} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, & \text{Modelo Normal.} \end{cases} \quad (2.16)$$

Por conta disso, será conveniente usarmos a seguinte ponderação

$$W_{it} = \frac{P_{it}^* Q_{it}^*}{P_{it} Q_{it}}, \quad (2.17)$$

onde $Q_{it}^* = 1 - P_{it}^*$. Para os modelos de 1 ou 2 parâmetros, $W_{it} = 1$, $i = 1, \dots, n$, $t = 1, \dots, T$. Multiplicando e dividindo a razão em (2.15) pelo termo $P_{it} Q_{it}$ e usando a ponderação (2.17) poderemos reescrever (2.15) como

$$\frac{\partial P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}. \quad (2.18)$$

Usando a notação

$$g_j^*(\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{\theta} | \mathbf{U}_{j..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \frac{P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta})}{P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}, \quad (2.19)$$

teremos que a função de estimação (2.9) pode ser escrita como

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (2.20)$$

Resta agora determinar as equações específicas para cada parâmetro do vetor $\boldsymbol{\zeta}_i$ da FRI adotada.

2.4.1 Aplicação ao Modelo Logístico de 3 Parâmetros (ML3)

Para obter as equações de estimação para os parâmetros do ML3, precisaremos das seguintes expressões:

$$\frac{\partial P_{it}}{\partial a_i} = D(1 - c_i)(\theta_t - b_i) P_{it}^* Q_{it}^* \quad (2.21)$$

$$\frac{\partial P_{it}}{\partial b_i} = -D a_i (1 - c_i) P_{it}^* Q_{it}^* \quad (2.22)$$

$$\frac{\partial P_{it}}{\partial c_i} = Q_{it}^* \quad (2.23)$$

Para o parâmetro de discriminação, temos de (2.20) e (2.21) que

$$\begin{aligned}
\frac{\partial \log L(\zeta, \eta)}{\partial a_i} &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial a_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\theta) d\theta \\
&= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) D(1 - c_i) (\theta_t - b_i) P_{it}^* Q_{it}^* \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\theta) d\theta \\
&= D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) (\theta_t - b_i) W_{it} \right\} g_j^*(\theta) d\theta.
\end{aligned}$$

Para o parâmetro de dificuldade, temos de (2.20) e (2.22) que

$$\begin{aligned}
\frac{\partial \log L(\zeta, \eta)}{\partial b_i} &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial b_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\theta) d\theta \\
&= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) (-1) D a_i (1 - c_i) P_{it}^* Q_{it}^* \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\theta) d\theta \\
&= -D a_i (1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \right\} g_j^*(\theta) d\theta.
\end{aligned}$$

Para o parâmetro de acerto ao acaso, temos de (2.20) e (2.23) que

$$\begin{aligned}
\frac{\partial \log L(\zeta, \eta)}{\partial c_i} &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial c_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\theta) d\theta \\
&= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) Q_{it}^* \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\theta) d\theta \\
&= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^*} \right\} g_j^*(\theta) d\theta.
\end{aligned}$$

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

$$a_i : \quad D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) (\theta_t - b_i) W_{it} \right\} g_j^*(\theta) d\theta = 0, \quad (2.24)$$

$$b_i : \quad -D a_i (1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \right\} g_j^*(\theta) d\theta = 0, \quad (2.25)$$

$$c_i : \quad \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^*} \right\} g_j^*(\theta) d\theta = 0. \quad (2.26)$$

Estas equações não possuem solução explícita e por isso precisaremos de algum método iterativo para a obtenção das estimativas de máxima verossimilhança dos parâmetros dos itens. Frequentemente, essa estimação é feita via algoritmo Newton-Raphson, para o qual se faz necessária a obtenção da derivada segunda da log-verossimilhança (Matriz Hessiana). Neste contexto, a estimação é realizada para todos os itens de uma só vez, o que envolve o tratamento de matrizes de ordem $3n$ para o ML3. Algumas alternativas podem ser propostas para facilitar a obtenção das estimativas. Por conta disso, na seção seguinte são apresentadas as expressões para as derivadas segundas da log-verossimilhança, bem como uma proposta que facilita a obtenção destas estimativas.

2.5 Implementação

Nesta seção apresentamos as derivadas segundas da log-verossimilhança com relação aos parâmetros dos itens. Os detalhes para obtenção dessas expressões podem ser encontrados na Seção A.1.

Sejam

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

e

$$\begin{aligned} \mathbf{H}_{ii(j)} &\equiv \frac{\partial^2 P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta}) / \partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}{P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})} \\ &= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{H}_{iit} - \sum_{t \in \tau_i} (U_{jit} - P_{it})^2 W_{it}^2 \mathbf{h}_{it} \mathbf{h}_{it}' + \right. \\ &\quad \left. + \left[\sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right] \left[\sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it}' \right] \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Para $l \neq i$,

$$\begin{aligned} \mathbf{H}_{il(j)} &\equiv \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})/\partial\zeta_i\partial\zeta'_l}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \\ &= \int_{\mathbb{R}^T} \left[\sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right] \left[\sum_{t \in \tau_i} (U_{jlt} - P_{lt}) W_{lt} \mathbf{h}'_{lt} \right] g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Ainda, por (2.18),

$$\begin{aligned} \mathbf{h}_{i(j)} &\equiv \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})/\partial\zeta_i}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\ \mathbf{f}(\boldsymbol{\zeta}_i) &\equiv \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial\zeta_i} = \sum_{j=1}^s r_j \mathbf{h}_{i(j)}, \end{aligned}$$

e

$$\mathbf{H}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_l) = \frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial\zeta_i \partial\zeta'_l} = \sum_{j=1}^s r_j \{ \mathbf{H}_{il(j)} - \mathbf{h}_{i(j)} \mathbf{h}'_{l(j)} \}. \quad (2.27)$$

Se $\widehat{\boldsymbol{\zeta}}^{(k)}$ é uma estimativa de $\boldsymbol{\zeta}$ na iteraçãõ k , entãõ na iteraçãõ $k + 1$ teremos que

$$\widehat{\boldsymbol{\zeta}}^{(k+1)} = \widehat{\boldsymbol{\zeta}}^{(k)} - [\mathbf{H}_I(\widehat{\boldsymbol{\zeta}}^{(k)})]^{-1} \mathbf{f}_I(\widehat{\boldsymbol{\zeta}}^{(k)}),$$

onde

$$\mathbf{f}_I(\boldsymbol{\zeta}) = \begin{pmatrix} \mathbf{f}(\boldsymbol{\zeta}_1) \\ \vdots \\ \mathbf{f}(\boldsymbol{\zeta}_n) \end{pmatrix}, \quad \mathbf{H}_I(\boldsymbol{\zeta}) = \begin{pmatrix} \mathbf{H}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_1) & \mathbf{H}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) & \cdots & \mathbf{H}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_n) \\ \mathbf{H}(\boldsymbol{\zeta}_2, \boldsymbol{\zeta}_1) & \mathbf{H}(\boldsymbol{\zeta}_2, \boldsymbol{\zeta}_2) & \cdots & \mathbf{H}(\boldsymbol{\zeta}_2, \boldsymbol{\zeta}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}(\boldsymbol{\zeta}_n, \boldsymbol{\zeta}_1) & \mathbf{H}(\boldsymbol{\zeta}_n, \boldsymbol{\zeta}_2) & \cdots & \mathbf{H}(\boldsymbol{\zeta}_n, \boldsymbol{\zeta}_n) \end{pmatrix}.$$

2.5.1 Usando Suposições Adicionais

Algumas suposições podem ser feitas de modo a facilitar o processo de estimação. Podemos notar que a estimação dos parâmetros dos itens, com os parâmetros populacionais conhecidos, envolve a inversão de matrizes de ordem $3n$. Para amenizar a exigência computacional, algumas modificações na estrutura de $\mathbf{H}_I(\boldsymbol{\zeta})$ podem ser feitas com base em uma suposição, denominada de *independência entre os itens*, de forma que

$$\mathbf{H}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_k) = \mathbf{0}, \quad \text{para } i \neq k. \quad (2.28)$$

Vale notar que esta suposição não é equivalente a independência local, que está relacionada aos indivíduos e declara que, fixada sua habilidade, um indivíduo responde aos itens de forma independente.

Com essa modificação, a estrutura de $\mathbf{H}_I(\boldsymbol{\zeta})$ fica

$$\mathbf{H}_I(\boldsymbol{\zeta}) = \begin{pmatrix} \mathbf{H}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}(\boldsymbol{\zeta}_2, \boldsymbol{\zeta}_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}(\boldsymbol{\zeta}_n, \boldsymbol{\zeta}_n) \end{pmatrix}. \quad (2.29)$$

Essa simplificação traz uma grande vantagem: passamos a inverter matrizes de ordem 3, ao invés de $3n$, no Algoritmo Newton-Raphson. O processo iterativo fica muito menos exigente, demandando um tempo bem menor.

Uma outra maneira de obtermos as estimativas dos parâmetros dos itens consiste aplicação do algoritmo EM (ver Dempster, Laird & Rubin, 1977). Essa abordagem será vista em detalhes no Capítulo 4.

Sobre a métrica

Após estimados os parâmetros dos itens, resta uma pergunta: *em que métrica os parâmetros dos itens foram estimados?* Como os parâmetros populacionais são conhecidos, então a métrica já está estabelecida, i.e., os valores dos parâmetros populacionais estão todos na mesma escala de medida e, portanto, são comparáveis. Logo, os parâmetros dos itens também serão comparáveis.

Completamos aqui o primeiro bloco para o processo de estimação, no que se refere à obtenção das equações de estimação para os parâmetros dos itens. No próximo capítulo tratamos do segundo bloco, relativo à obtenção das equações de estimação dos parâmetros populacionais.

Modelo Longitudinal para Uma Única População: Estimação dos Parâmetros Populacionais

Neste capítulo, tratamos da estimação dos parâmetros populacionais considerando os parâmetros dos itens conhecidos e o conjunto de dados completos. Essa situação ocorre se os testes envolverem itens já calibrados, de forma que não há necessidade na reestimação desses parâmetros. Entretanto, essa situação também pode ser vista como um segundo bloco para a estimação conjunta dos parâmetros dos itens e parâmetros populacionais.

Nossos passos serão, inicialmente, a obtenção das equações de estimação para $\boldsymbol{\eta}$ sem especificar a $fdp\ g(\boldsymbol{\theta}|\boldsymbol{\eta})$. Feito isso, exploraremos o caso em que $\boldsymbol{\theta}$ tem distribuição Normal Multivariada com parâmetros $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, onde $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_T)'$ é o vetor de médias nos T testes e $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j \leq T}$, com $\sigma_{tt} \equiv \sigma_t^2$, é a matriz de covariâncias. Consideraremos 5 estruturas para $\boldsymbol{\Sigma}$. A saber, os modelos Diagonal, Uniforme, Bandas de ordem 1, Auto-Regressivo de ordem 1 e Hankel. Para o primeiro caso teremos que $\boldsymbol{\eta} = (\mu_1, \mu_2, \mu_3, \dots, \mu_T, \sigma_{11}, \sigma_{22}, \dots, \sigma_{TT})'$. Para os 3 seguintes teremos que $\boldsymbol{\eta} = (\mu_1, \mu_2, \mu_3, \dots, \mu_T, \sigma^2, \rho)'$, onde ρ representa a correlação entre as habilidades em qualquer par de testes. Para o último caso teremos $\boldsymbol{\eta} = (\mu_1, \mu_2, \mu_3, \dots, \mu_T, \sigma_{12}, \sigma_{11}, \sigma_{22}, \dots, \sigma_{TT})'$, onde σ_{12} representa a covariância entre as habilidades em qualquer par de testes.

3.1 Equação de Estimação: Caso Geral

As equações de estimação para os parâmetros populacionais são dadas por

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \mathbf{0}. \quad (3.1)$$

Para especificarmos essas equações, notemos que

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \sum_{j=1}^s r_j \frac{1}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}. \quad (3.2)$$

Mas

$$\begin{aligned}\frac{\partial P(\mathbf{U}_{j..}|\zeta, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} &= \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \zeta) \left(\frac{\partial}{\partial \boldsymbol{\eta}} g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right) d\boldsymbol{\theta}.\end{aligned}\quad (3.3)$$

Segue de (3.2) que a equação de estimação para $\boldsymbol{\eta}$ é

$$\begin{aligned}\frac{\partial \log L(\zeta, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} &= \sum_{j=1}^s r_j \frac{1}{P(\mathbf{U}_{j..}|\zeta, \boldsymbol{\eta})} \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right) d\boldsymbol{\theta} \\ &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right) g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbf{0}.\end{aligned}\quad (3.4)$$

Várias distribuições podem ser propostas para as habilidades. Uma das distribuições de maior interesse é a Normal, para a qual obteremos as equações de estimação a seguir.

3.2 Habilidades Normalmente Distribuídas

No caso em que $\boldsymbol{\theta}$ tem distribuição Normal T -variada com vetor de médias $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_T)'$ e matriz de covariâncias $\boldsymbol{\Sigma}$, teremos que

$$\log g(\boldsymbol{\theta}|\boldsymbol{\eta}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}).$$

Com isso, segue que

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

Em particular, considerando $\mathbf{V} = \boldsymbol{\Sigma}^{-1} = (v_{ij})$, temos que

$$\begin{aligned}\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_t} &= -\frac{1}{2} \frac{\partial}{\partial \mu_t} [(\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{V} (\boldsymbol{\theta} - \boldsymbol{\mu})] \\ &= \mathbf{v}'_t (\boldsymbol{\theta} - \boldsymbol{\mu}), \quad t = 1, \dots, T,\end{aligned}\quad (3.5)$$

onde \mathbf{v}_t é a t -ésima coluna de \mathbf{V} . Em termos matriciais, podemos escrever as equações (3.5) como

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu}} = \begin{pmatrix} \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_1} \\ \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_2} \\ \vdots \\ \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_T} \end{pmatrix} = \begin{pmatrix} \mathbf{v}'_1 \\ \mathbf{v}'_2 \\ \vdots \\ \mathbf{v}'_T \end{pmatrix} (\boldsymbol{\theta} - \boldsymbol{\mu}) = \mathbf{V}(\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (3.6)$$

Usando a notação

$$\tilde{\boldsymbol{\mu}}_j = \int_{\mathbb{R}^T} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad \text{e} \quad \bar{\boldsymbol{\mu}} = \frac{1}{N} \sum_{j=1}^s r_j \tilde{\boldsymbol{\mu}}_j \quad (3.7)$$

teremos, de (3.4) e (3.6), que

$$\begin{aligned} \mathbf{0} &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \mathbf{V}(\boldsymbol{\theta} - \boldsymbol{\mu}) g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \sum_{j=1}^s r_j \mathbf{V} \left\{ \int_{\mathbb{R}^T} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} - \boldsymbol{\mu} \int_{\mathbb{R}^T} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \right\} \\ &= \sum_{j=1}^s r_j \mathbf{V} \{ \tilde{\boldsymbol{\mu}}_j - \boldsymbol{\mu} \} \\ &= \mathbf{V} \left\{ \sum_{j=1}^s r_j \tilde{\boldsymbol{\mu}}_j - N \boldsymbol{\mu} \right\} \\ &= N \mathbf{V} \{ \bar{\boldsymbol{\mu}} - \boldsymbol{\mu} \}. \end{aligned} \quad (3.8)$$

Para fins de estimação de $\boldsymbol{\eta}$, ainda são necessárias as equações relativas aos parâmetros da estrutura de covariância adotada. De forma geral, quando a matriz $\boldsymbol{\Sigma}$ é não-estruturada, isto é, tem $T(T+1)/2$ parâmetros, vem

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \sigma_{ij}} = -\frac{1}{2} \frac{\partial \log |\boldsymbol{\Sigma}|}{\partial \sigma_{ij}} - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \left[\frac{\partial}{\partial \sigma_{ij}} \mathbf{V} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (3.9)$$

Considerando $\delta_{ts} = \partial \log |\boldsymbol{\Sigma}| / \partial \sigma_{ts}$ e $\mathbf{V}_{ts}^{(1)} = \partial \mathbf{V} / \partial \sigma_{ts}$, e usando a decomposição $\boldsymbol{\theta} - \boldsymbol{\mu} = (\boldsymbol{\theta} - \bar{\boldsymbol{\mu}}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})$ para a equação de estimação para σ_{ts} temos que

$$\begin{aligned}
0 &= -\frac{1}{2} \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \delta_{ts} + (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{V}_{ts}^{(1)} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= -\frac{1}{2} \left\{ N\delta_{ts} + \sum_{j=1}^s r_j \int_{\mathbb{R}^T} (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{V}_{ts}^{(1)} (\boldsymbol{\theta} - \boldsymbol{\mu}) g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \right\} \\
&= -\frac{N}{2} \left\{ \delta_{ts} + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})' \mathbf{V}_{ts}^{(1)} (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}) + \int_{\mathbb{R}^T} (\boldsymbol{\theta} - \bar{\boldsymbol{\mu}})' \mathbf{V}_{ts}^{(1)} (\boldsymbol{\theta} - \bar{\boldsymbol{\mu}}) \bar{g}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \right\}, \quad (3.10)
\end{aligned}$$

onde $\bar{g}^*(\boldsymbol{\theta}) = N^{-1} \sum_{j=1}^s r_j g_j^*(\boldsymbol{\theta})$. Pelas Equações (3.8) e (3.10), os EMV de $\boldsymbol{\mu}$ e σ_{ts} são obtidos resolvendo-se as equações

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}} \quad (3.11)$$

$$\int_{\mathbb{R}^T} (\boldsymbol{\theta} - \bar{\boldsymbol{\mu}})' \mathbf{V}_{ts}^{(1)} (\boldsymbol{\theta} - \bar{\boldsymbol{\mu}}) \bar{g}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = -\delta_{ts}. \quad (3.12)$$

Portanto, para chegarmos às equações de estimação para $\boldsymbol{\eta}$ basta que obtenhamos as expressões das derivadas primeiras de $\log |\boldsymbol{\Sigma}|$ e de $\mathbf{V} = \boldsymbol{\Sigma}^{-1}$. Basicamente, será isso que faremos a seguir para as diversas estruturas de covariâncias de interesse.

3.3 Matriz de Covariância Diagonal

A estrutura mais simples que poderíamos considerar seria $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$, na qual todas as variâncias são iguais e as correlações nulas. Uma segunda alternativa seria a matriz Diagonal, a qual é apropriada em situações onde as correlações podem ser consideradas nulas, mas as variâncias podem diferir. Essa estrutura é dada por

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{pmatrix} \equiv \mathbf{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_T^2\}.$$

Neste caso temos que

$$\begin{aligned}
|\boldsymbol{\Sigma}| &= \prod_{t=1}^T \sigma_t^2, \\
\mathbf{V} &= \boldsymbol{\Sigma}^{-1} = \mathbf{diag}\{1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_T^2\}.
\end{aligned}$$

Logo, para $t, i, j = 1, \dots, T$,

$$\frac{\partial \log |\Sigma|}{\partial \sigma_t^2} = \frac{1}{\sigma_t^2}, \quad (3.13)$$

$$\mathbf{V}_t^{(1)} = \left(\frac{\partial v_{ij}}{\partial \sigma_t^2} \right),$$

onde

$$\frac{\partial v_{ij}}{\partial \sigma_t^2} = \begin{cases} -1/\sigma_t^4 & \text{se } i = j = t ; \\ 0 & \text{caso contrário.} \end{cases} \quad (3.14)$$

Para chegarmos às equações de estimação para $\boldsymbol{\eta}$, basta substituímos as expressões (3.13) e (3.14) em (3.9) e o resultado em (3.4).

No sentido de apresentar uma forma mais simplificada para essas equações, notemos que a distribuição conjunta das habilidades nos T tempos é o produto das distribuições marginais, ou seja,

$$g(\boldsymbol{\theta}|\boldsymbol{\eta}) = \prod_{t=1}^T g_t(\theta_t|\boldsymbol{\eta}_t), \quad (3.15)$$

onde $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, \dots, \boldsymbol{\eta}'_T)'$, com $\boldsymbol{\eta}_t = (\mu_t, \sigma_t^2)'$. Segue que

$$\begin{aligned} P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) &= \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} \prod_{t=1}^T P(\mathbf{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) \prod_{t=1}^T g_t(\theta_t|\boldsymbol{\eta}_t) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} \prod_{t=1}^T \{P(\mathbf{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) g_t(\theta_t|\boldsymbol{\eta}_t)\} d\boldsymbol{\theta} \\ &= \prod_{t=1}^T \left\{ \int_{\mathbb{R}} P(\mathbf{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) g_t(\theta_t|\boldsymbol{\eta}_t) d\theta_t \right\} \\ &= \prod_{t=1}^T P(\mathbf{U}_{j.t}|\boldsymbol{\zeta}, \boldsymbol{\eta}_t). \end{aligned}$$

Logo, por (2.3), (3.15) e (3.16)

$$\begin{aligned} g_j^*(\boldsymbol{\theta}) &= \frac{\left[\prod_{t=1}^T P(\mathbf{U}_{j.t} | \theta_t, \boldsymbol{\eta}_t) \right] \left[\prod_{t=1}^T g_t(\theta_t | \boldsymbol{\eta}_t) \right]}{\prod_{t=1}^T P(\mathbf{U}_{j.t} | \boldsymbol{\zeta}, \boldsymbol{\eta}_t)} \\ &= \prod_{t=1}^T g_{jt}^*(\theta_t), \end{aligned}$$

onde

$$g_{jt}^*(\theta_t) = \frac{P(\mathbf{U}_{j.t} | \theta_t, \boldsymbol{\zeta}) g_t(\theta_t | \boldsymbol{\eta}_t)}{P(\mathbf{U}_{j.t} | \boldsymbol{\zeta}, \boldsymbol{\eta}_t)}.$$

Ainda,

$$\frac{\partial \log g(\boldsymbol{\theta} | \boldsymbol{\eta})}{\partial \boldsymbol{\eta}_t} = \frac{\partial \log g_t(\theta_t | \boldsymbol{\eta}_t)}{\partial \boldsymbol{\eta}_t} = \begin{pmatrix} \frac{1}{\sigma_t^2} (\theta_t - \mu_t) \\ -\frac{1}{2\sigma_t^4} (\sigma_t^2 - (\theta_t - \mu_t)^2) \end{pmatrix}.$$

Substituindo-se os resultados acima em (3.4) obtém-se

$$\begin{aligned} \mu_t : \quad & \frac{1}{\sigma_t^2} \sum_{j=1}^{s_t} r_{jt} \frac{1}{P(\mathbf{U}_{j.t} | \boldsymbol{\zeta}, \boldsymbol{\eta}_t)} \int_{\mathbb{R}} P(\mathbf{U}_{j.t} | \theta, \boldsymbol{\zeta}) (\theta - \mu_t) g(\theta | \boldsymbol{\eta}_t) d\theta = 0, \\ \sigma_t^2 : \quad & -\frac{1}{2\sigma_t^4} \sum_{j=1}^{s_t} r_{jt} \frac{1}{P(\mathbf{U}_{j.t} | \boldsymbol{\zeta}, \boldsymbol{\eta}_t)} \int_{\mathbb{R}} P(\mathbf{U}_{j.t} | \theta, \boldsymbol{\zeta}) (\sigma_t^2 - (\theta - \mu_t)^2) g(\theta | \boldsymbol{\eta}_t) d\theta = 0. \end{aligned}$$

Estas equações são exatamente as mesmas equações de estimação obtidas por Bock & Zimowski (1997), onde os tempos correspondem aos grupos .

Na situação em que os parâmetros dos itens são desconhecidos, podem-se utilizar argumentos similares aos utilizados acima para obter as equações de estimação para esses parâmetros. Essas equações estão apresentadas abaixo e são exatamente as mesmas obtidas por Bock & Zimowski.

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} &= \sum_{t \in \tau_i} \sum_{j=1}^s r_j \int_{\mathbb{R}} \left\{ (U_{jit} - P_i) \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_i}{P_i^* Q_i^*} \right\} g_{jt}^*(\theta) d\theta \\ &= \sum_{t \in \tau_i} \sum_{j=1}^{s_t} r_{jt} \int_{\mathbb{R}} \left\{ (U_{jit} - P_i) \left(\frac{\partial P_i}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_i}{P_i^* Q_i^*} \right\} g_{jt}^*(\theta) d\theta, \quad (3.16) \end{aligned}$$

na qual o conjunto de padrões de resposta foi reestruturado de forma a termos um conjunto para cada teste; r_{jt} representa o número de ocorrências do padrão de resposta j no teste t e s_t o número de padrões de resposta no teste t com $r_{jt} > 0$. Além disso, a substituição de θ_t por θ permite simplificar as notações $P_{it}, P_{it}^*, Q_{it}^*$ e W_{it} para P_i, P_i^*, Q_i^* e W_i , respectivamente.

3.4 Matriz de Covariância Uniforme

Essa estrutura de covariância tem sido bastante explorada em estudos longitudinais (ver Singer & Andrade, 2000). Com ela, temos que as variâncias são iguais a σ^2 nas T condições de avaliação, e que a correlação $\rho \in [-1, 1]$ entre as habilidades em qualquer par (θ_t, θ_s) é sempre a mesma. Sua estrutura é dada por

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

Esta matriz pode ser escrita como $\Sigma = \sigma^2(1 - \rho)\mathbf{I}_T + \sigma^2\rho\mathbf{J}_T$, onde \mathbf{I}_T é a matriz identidade de ordem T , $\mathbf{J}_T = \mathbf{1}_T\mathbf{1}'_T$, e $\mathbf{1}_T = (1, \dots, 1)'$ é o vetor $T \times 1$ de elementos unitários. Para esta matriz temos que

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

Usando a notação

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

teremos que

$$|\Sigma| = \sigma^{2T} \alpha^{T-1} \beta, \tag{3.17}$$

$$\mathbf{V} = \frac{1}{\alpha\sigma^2} \left(\mathbf{I}_T - \frac{\rho}{\beta} \mathbf{J}_T \right). \tag{3.18}$$

Segue de (3.17) que

$$\frac{\partial \log |\Sigma|}{\partial \rho} = -(T - 1) \left[\frac{1}{\alpha} - \frac{1}{\beta} \right]. \tag{3.19}$$

$$\frac{\partial \mathbf{V}}{\partial \rho} = \frac{1}{\sigma^2} \left\{ \frac{1}{\alpha^2} \mathbf{I}_T - \left(\frac{1}{\alpha\beta} - \frac{\rho c}{(\alpha\beta)^2} \right) \mathbf{J}_T \right\}. \quad (3.20)$$

De forma geral, para as estruturas Uniforme, Bandas e Auto-Regressiva de ordem 1, temos que $\Sigma = \sigma^2 \mathbf{R}$, onde \mathbf{R} é uma matriz de correlação. Com isso,

$$\frac{\partial \log |\Sigma|}{\partial \sigma^2} = \frac{\partial [\log \sigma^{2T}]}{\partial \sigma^2} = \frac{T}{\sigma^2}. \quad (3.21)$$

e

$$\frac{\partial \mathbf{V}}{\partial \sigma^2} = -\sigma^{-2} \mathbf{V}. \quad (3.22)$$

Para chegarmos às equações de estimação para $\boldsymbol{\eta}$, basta substituímos as expressões (3.19) a (3.22) em (3.9) e o resultado em (3.4).

3.5 Matriz de Covariância de Bandas

Essa estrutura de covariância, também chamada de TOEPLITZ, tem uma diferença importante com relação à estrutura uniforme: apenas as habilidades nos tempos imediatamente anterior e posterior a um particular tempo são dependentes entre si, tornando-a mais próxima do indicado em muitas situações práticas, nas quais espera-se que as correlações diminuam quando o tempo decorrido entre as observações aumenta. A forma dessa estrutura de covariância é dada por

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Para esta matriz temos que

$$\begin{aligned} |\Sigma| &= \sigma^{2T} \prod_{t=1}^T \left\{ 1 + 2\rho \cos \left(\frac{t\pi}{T+1} \right) \right\}, \\ \Sigma^{-1} &= \mathbf{V} = (v_{ij})_{i,j \leq T}, \text{ com} \end{aligned}$$

$$v_{ij} = \frac{(1 - \beta^{2(T-m_{ij}+1)})(\beta^{j+i+1} - \beta^{|j-i|+1})}{\sigma^2 \rho (1 - \beta^2)(1 - \beta^{2(T+1)}),}$$

onde $m_{ij} = \max\{i, j\}$ e $\beta = \frac{1}{2\rho}(\sqrt{1 - 4\rho^2} - 1)$, ou equivalentemente,

$$\rho = \frac{-\beta}{1 + \beta^2}. \quad (3.23)$$

Para Σ ser positiva-definida, devemos ter $\rho \in [0, \frac{1}{2}]$ (ver Graybill, 1969), e nessa situação teremos que $\beta \in [-1, 0]$ e

$$\frac{\partial}{\partial \rho} \log |\Sigma| = \sum_{t=1}^T \left\{ \rho + \frac{1}{2} \cos^{-1} \left(\frac{t\pi}{T+1} \right) \right\}^{-1}. \quad (3.24)$$

Considerando $v_{ij} = \gamma_{1ij}\gamma_{2ij}/(\rho\sigma^2\gamma_3)$, onde

$$\gamma_{1ij} = 1 - \beta^{2(T-m_{ij}+1)}, \quad \gamma_{2ij} = \beta^{j+i+1} - \beta^{|j-i|+1} \quad \text{e} \quad \gamma_3 = (1 - \beta^2)(1 - \beta^{2(T-1)}),$$

temos que

$$\begin{aligned} \beta^{(1)} &= \frac{\partial \beta}{\partial \rho} = -\frac{(1 + \beta^2)^2}{1 - \beta^2} \\ \gamma'_{1ij} &= \frac{\partial \gamma_{1ij}}{\partial \rho} = -2(T - m_{ij} + 1)\beta^{2(T-m_{ij}+1)}\beta^{(1)} \\ \gamma'_{2ij} &= \frac{\partial \gamma_{2ij}}{\partial \rho} = \{(j + i + 1)\beta^{j+i} - (|j - i| + 1)\beta^{|j-i|}\} \beta^{(1)} \\ \gamma'_3 &= \frac{\partial \gamma_3}{\partial \rho} = -2\beta\beta^{(1)} \{1 - \beta^{2T}[(T + 2)\beta^2 - (T + 1)]\}. \end{aligned}$$

e

$$\frac{\partial v_{ij}}{\partial \rho} = \frac{1}{\sigma^2(\rho\gamma_3)^2} [\rho\gamma_3(\gamma'_{1ij}\gamma_{2ij} + \gamma_{1ij}\gamma'_{2ij}) - \gamma_{1ij}\gamma_{2ij}(\gamma_3 + \rho\gamma'_3)] \equiv \frac{\gamma_{4ij}}{\sigma^2(\rho\gamma_3)^2}. \quad (3.25)$$

As derivadas de $\log |\Sigma|$ e \mathbf{V} com relação a σ^2 são dadas por (3.21) e (3.22), respectivamente. Para chegarmos às equações de estimação para $\boldsymbol{\eta}$, basta substituímos as expressões (3.24) e (3.25), bem como (3.21) e (3.22), em (3.9) e o resultado em (3.4).

3.6 Matriz de Covariância AR(1)

Essa estrutura de covariância é bastante importante. Assim como as duas últimas estruturas, ela também considera que as variâncias não se alteram com o decorrer do tempo, mas assume que as correlações diminuem monotonicamente, diferentemente da estrutura de Bandas em que estas são nulas quando o tempo decorrido entre os testes é maior que uma unidade. Isso torna essa estrutura mais realista do que as anteriores em muitas situações práticas. Para experimentos balanceados, nos quais os intervalos de tempo entre os testes são sempre os mesmos, a forma dessa estrutura de covariância é dada por

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}.$$

Esta matriz pode ser escrita como $\Sigma = (\sigma_{ij})_{i,j \leq T}$ com $\sigma_{ij} = \sigma^2 \rho^{|i-j|}$. Usando a notação

$$\alpha = \alpha(\rho) = 1 - \rho^2, \quad \beta = \beta(\rho) = 1 + \rho^2,$$

teremos que o determinante e inversa de Σ são, respectivamente

$$\begin{aligned} |\Sigma| &= \sigma^{2T} \alpha^{T-1}, \\ \mathbf{V} &= \Sigma^{-1} = (\sigma^{-2} v_{ij})_{i,j \leq T}, \end{aligned}$$

com

$$v_{ij} = \begin{cases} 1/\alpha & \text{se } i = j \text{ e } i \in \{1, T\} \\ \beta/\alpha & \text{se } i = j \text{ e } i \in \{2, 3, \dots, T-1\} \\ -\rho/\alpha & \text{se } j \in \{i-1, i+1\} \\ 0 & \text{c.c.} \end{cases}$$

Segue que

$$\frac{\partial \log |\Sigma|}{\partial \rho} = \frac{-2\rho(T-1)}{\alpha}, \quad (3.26)$$

e

$$\frac{\partial v_{ij}}{\partial \rho} = \begin{cases} \frac{2\rho}{\alpha^2} & \text{se } i = j \text{ e } i \in \{1, T\} \\ \frac{4\rho}{\alpha^2} & \text{se } i = j \text{ e } i \in \{2, 3, \dots, T-1\} \\ -\frac{\beta}{\alpha^2} & \text{se } j \in \{i-1, i+1\} \\ 0 & \text{c.c.} \end{cases} \quad (3.27)$$

As derivadas de $\log |\Sigma|$ e \mathbf{V} com relação a σ^2 são dadas por (3.21) e (3.22), respectivamente. Para chegarmos às equações de estimação para $\boldsymbol{\eta}$, basta substituímos as expressões (3.26) e (3.27), bem como (3.21) e (3.22), em (3.9) e o resultado em (3.4).

3.7 Matriz de Covariância de Hankel

Essa estrutura de covariância é bastante conveniente. Ela permite que as variâncias nos T tempos sejam diferentes, possibilitando que as correlações entre cada par (θ_t, θ_s) sejam diferentes. Nenhuma das estruturas anteriores tinha essa propriedade. Como consequência, o número de parâmetros desta estrutura é maior que o das anteriores. Sua forma é dada por

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{12} & \cdots & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 & \sigma_{12} & \cdots & \sigma_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{12} & \sigma_{12} & \sigma_{12} & \cdots & \sigma_T^2 \end{pmatrix}.$$

Considerando $\mathbf{D}_T = \mathbf{diag}(d_1, d_2, \dots, d_T)$, com $d_t = \sigma_t^2 - \sigma_{12}$, podemos reescrever $\Sigma = \sigma_{12} \mathbf{J}_T + \mathbf{D}_T$. Ainda, se

$$\alpha = 1 + \sigma_{12} \sum_{t=1}^T \frac{1}{d_t} \quad \text{e} \quad \beta = \prod_{t=1}^T d_t,$$

então,

$$\begin{aligned} |\Sigma| &= \alpha\beta, \\ \mathbf{V} &= \Sigma^{-1} = \mathbf{F}_T + \gamma \mathbf{E}_T, \end{aligned}$$

onde $\gamma = -\sigma_{12}/\alpha$, $\mathbf{F}_T = \mathbf{D}_T^{-1}$ e $\mathbf{E} = (e_{ij})$ com $e_{ij} = 1/(d_i d_j)$.

Sejam

$$S_k = \sum_{t=1}^T \frac{1}{d_t^k} \quad \text{e} \quad \alpha_k = S_k + \sigma_{12} S_{k+1} \quad \text{para } k \geq 0,$$

com $S_0 = 1$. Segue que

$$\frac{\partial}{\partial \sigma_{12}} \log |\boldsymbol{\Sigma}| = \frac{\alpha_1}{\alpha} - S_1; \quad (3.28)$$

$$\frac{\partial}{\partial \sigma_{tt}} \log |\boldsymbol{\Sigma}| = \frac{\gamma}{d_t^2} + \frac{1}{d_t}, \quad t = 1, \dots, T. \quad (3.29)$$

As derivadas de \mathbf{V} com relação a σ_{12} e σ_{tt} são, respectivamente,

$$\mathbf{V}_{12}^{(1)} = \frac{\partial \mathbf{V}}{\partial \sigma_{12}} = \mathbf{F}_{12}^{(1)} + \gamma_{12}^{(1)} \mathbf{E} + \gamma \mathbf{E}_{12}^{(1)}; \quad (3.30)$$

$$\mathbf{V}_{tt}^{(1)} = \frac{\partial \mathbf{V}}{\partial \sigma_{tt}} = \mathbf{F}_{tt}^{(1)} + \gamma_{tt}^{(1)} \mathbf{E} + \gamma \mathbf{E}_{tt}^{(1)}, \quad t = 1, \dots, T. \quad (3.31)$$

onde

$$\gamma_{12}^{(1)} = \delta/\alpha^2, \quad \gamma_{tt}^{(1)} = -\gamma^2/d_t^2 \quad \text{e} \quad \delta = \sigma_{12}\alpha_1 - \alpha,$$

As derivadas de $\mathbf{F} = (f_{ts})$ são

$$f_{ts,12}^{(1)} = \begin{cases} 1/d_i^2 & \text{se } t = s = i \\ 0 & \text{c.c.} \end{cases}$$

$$f_{ts,ii}^{(1)} = \begin{cases} -1/d_i^2 & \text{se } t = s = i \\ 0 & \text{c.c.} \end{cases}$$

As derivadas de $\mathbf{E} = (e_{ij})$ são

$$e_{ij,12}^{(1)} = \frac{(d_i + d_j)}{(d_i d_j)^2};$$

$$e_{ij,tt}^{(1)} = \begin{cases} -1/(d_i^2 d_j) & \text{se } t = i, i \neq j \\ -2/d_i^3 & \text{se } t = i = j \\ 0 & \text{c.c.} \end{cases}$$

Para chegarmos às equações de estimação para $\boldsymbol{\eta}$, basta substituímos as expressões (3.28) e (3.30) em (3.9) e o resultado em (3.4).

3.8 Comentários Adicionais

- Outras estruturas de covariância poderiam ser sugeridas, como por exemplo a não-estruturada, que é composta por $T(T+1)/2$ parâmetros. Neste caso, os cálculos são um pouco mais complicados, pois não existem formas fechadas tanto para a sua inversa quanto para o seu determinante.

- Para qualquer das estruturas de covariância adotadas, as equações de estimação obtidas não apresentam solução explícita e por isso precisaremos de algum método iterativo para a obtenção dessas estimativas. Na Seção A.2 apresenta-se o método Newton-Raphson.

Sobre a métrica

Considerando que os parâmetros populacionais já foram estimados, resta saber em que métrica eles foram obtidos. Estamos tratando do caso em que os parâmetros dos itens são conhecidos, isto é, já foram estimados, e o conhecimento dos parâmetros dos itens implica o conhecimento da métrica em que eles foram obtidos. Por consequência, os parâmetros populacionais serão condizentes com a métrica dos parâmetros dos itens.

Completamos aqui o segundo bloco para o processo de estimação conjunta, no que se refere à obtenção das equações de estimação para os parâmetros populacionais. No capítulo seguinte os dois blocos serão unidos para o tratamento da estimação conjunta: parâmetros dos itens e populacionais.

Modelo Longitudinal para Uma Única População: Estimação Conjunta - Parâmetros dos Itens e Populacionais

Neste capítulo tratamos do caso mais freqüente, em que precisamos estimar tanto os parâmetros dos itens quanto os populacionais. Basicamente, o procedimento seria considerar como equações de estimação o conjunto obtido pela união das equações finais dos Capítulos 2 e 3. Entretanto, na estimação conjunta há 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 não-identificável. A forma mais usual de eliminar esse problema é definir uma *População Referência* e a *métrica* associada à ela. Além disso, com um número maior de parâmetros a estimar, algumas alternativas podem ser propostas de forma a facilitar a implementação do processo de estimação.

População Referência e Métrica

Como comentado na Seção 1.3, toda variável latente necessita do arbítrio de uma origem e uma unidade de medida, pois em caso contrário, qualquer sistema de equações para estimação dos parâmetros será não-identificável. Isso pode ser visto facilmente se considerarmos a seguinte transformação linear: $\theta_{jt}^* = \alpha\theta_{jt} + \beta$, $b_i^* = \alpha b_i + \beta$ e $a_i^* = a_i/\alpha$, onde α e β são constantes reais com $\alpha > 0$. Com isso, obtemos $P(U_{jit} = 1|\theta_{jt}^*, \zeta_i^*) = P(U_{jit} = 1|\theta_{jt}, \zeta_i)$.

Neste trabalho adotaremos o tempo 1 como referência, fazendo

$$\mu_1 = 0 \quad \text{e} \quad \sigma_1^2 = 1. \quad (4.1)$$

Com estes parâmetros fixados, as expressões do Capítulo 3 devem ser adaptadas de modo a não incluir a estimação de μ_1 e σ_1^2 . O vetor de parâmetros de médias a estimar é $\boldsymbol{\mu}_{(1)} = (\mu_2, \dots, \mu_T)'$. Em particular, para o caso em que $\boldsymbol{\theta} \sim N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a expressão (3.6) passa a ser escrita como

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu}_{(1)}} = \mathbf{V}_{(1)}(\boldsymbol{\theta} - \boldsymbol{\mu}), \quad (4.2)$$

onde $\mathbf{V}_{(1)}$ é a matriz \mathbf{V} sem a primeira linha. As expressões referentes à estimação de σ_1^2 não se fazem mais necessárias. Em particular, para as estruturas Uniforme, Bandas e AR(1), teremos $\sigma^2 = 1$.

Considerados esses detalhes, as equações de estimação para o conjunto restante de parâmetros populacionais são exatamente as mesmas apresentadas no Capítulo 3, que devem ser unidas às equações de estimação dos parâmetros dos itens para formar o conjunto integral das equações de estimação, ou seja

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbf{0}, \quad i = 1, \dots, n, \\ \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} &= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right) g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbf{0}. \end{aligned} \quad (4.3)$$

As derivadas segundas da log-verossimilhança com relação aos parâmetros dos itens estão dadas na Seção A.1, com relação aos parâmetros populacionais na Seção A.2 e com relação aos parâmetros dos itens e populacionais na Seção A.3.

4.1 Implementação

Na estimação conjunta, temos um número maior de parâmetros a estimar. Dessa forma, se faz necessário o uso de artifícios para facilitar o processo de estimação. Algumas suposições podem ser feitas nesse sentido. A primeira delas é a de *independência entre itens*, dada por (2.28), e a segunda é a de *independência entre itens e população*, cujo objetivo é tornar nula a derivada segunda da log-verossimilhança com relação a $\boldsymbol{\zeta}_i$ e $\boldsymbol{\eta}$, ou seja,

$$\mathbf{H}_{iP}(\boldsymbol{\zeta}_i, \boldsymbol{\eta}) = \mathbf{0}, \quad \text{para } i = 1, \dots, n. \quad (4.4)$$

Considerando $\boldsymbol{\psi} = (\boldsymbol{\zeta}', \boldsymbol{\eta}')'$ o conjunto integral de parâmetros a estimar, e $\mathbf{H}(\boldsymbol{\psi})$ a matriz Hessiana relativa a esses parâmetros, por (2.28) e (4.4) a estrutura de $\mathbf{H}(\boldsymbol{\psi})$ fica

$$\mathbf{H}_{IP}(\boldsymbol{\psi}) = \begin{pmatrix} \mathbf{H}_I(\boldsymbol{\zeta}) & \mathbf{0} \\ \mathbf{0}' & \mathbf{H}_P(\boldsymbol{\eta}) \end{pmatrix},$$

onde $\mathbf{H}_P(\boldsymbol{\eta})$ é a matriz das derivadas segundas relativas aos parâmetros populacionais.

Novamente, essa estrutura envolve o tratamento de matrizes de dimensões muito menores que as originais, simplificando bastante o processo de estimação.

Um outro processo de determinação de estimativas de máxima verossimilhança é o algoritmo EM (ver Dempster, Laird & Rubin, 1977). Com ele, não são necessárias as suposições de independência entre itens e independência entre itens e população.

4.1.1 Aplicação do Algoritmo EM

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

Passo E: Calcular $E[\log f(\mathbf{U}_{\dots}, \boldsymbol{\theta} | \boldsymbol{\psi}) | \mathbf{U}_{\dots}, \hat{\boldsymbol{\psi}}^{(k)}]$

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

No passo M a maximização pode ser feita pelo algoritmo Newton-Raphson ou “Scoring” de Fisher, por exemplo.

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

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

Para descrever brevemente o algoritmo EM aplicado à TRI, consideremos

$$f(\boldsymbol{\theta}) = \sum_{j=1}^s r_j g_j^*(\boldsymbol{\theta}) \quad \text{e} \quad r_{it}(\boldsymbol{\theta}) = \sum_{j=1}^s r_j U_{jit} g_j^*(\boldsymbol{\theta}), \quad (4.5)$$

que representam, respectivamente, o número de indivíduos em uma população de tamanho N que têm vetor de habilidades $\boldsymbol{\theta}$ e o número de tais indivíduos que respondem *corretamente* ao item i no teste t . Os conjuntos $\mathbf{f} = (f(\boldsymbol{\theta}))$ e $\mathbf{r} = (r_{it}(\boldsymbol{\theta}))$ são a representação de

$\boldsymbol{\theta}$ e compõem as quantidades não-observadas do algoritmo EM. Com isso, da expressão (2.20) temos que as novas equações de estimação para os parâmetros dos itens são dadas por

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} &= \int_{\mathbb{R}^T} \sum_{t \in \tau_i} \left\{ \sum_{j=1}^s r_j [(U_{jit} - P_{it}) g_j^*(\boldsymbol{\theta})] \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} \sum_{t \in \tau_i} \left\{ \left[\sum_{j=1}^s r_j U_{jit} g_j^*(\boldsymbol{\theta}) - \sum_{j=1}^s r_j P_{it} g_j^*(\boldsymbol{\theta}) \right] \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} \sum_{t \in \tau_i} \left\{ [r_{it}(\boldsymbol{\theta}) - f(\boldsymbol{\theta}) P_{it}] \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} d\boldsymbol{\theta}. \end{aligned} \quad (4.6)$$

As equações de estimação para os parâmetros populacionais são dadas por

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \int_{\mathbb{R}^T} \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta} | \boldsymbol{\eta}) \right) f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbf{0}. \quad (4.7)$$

Na prática, as integrais são aproximadas por meio de um dos muitos métodos disponíveis na literatura. Na TRI, tem sido freqüente a aplicação do método *Hermite-Gauss* (ver Stroud & Secrest, 1966), usualmente denominado de *método de quadratura gaussiana*, segundo o qual se considera a aproximação com base em um conjunto de Q pontos $\boldsymbol{\theta}_l$ com pesos A_l , $l = 1, \dots, Q$. Com base nisso, os passos E e M podem ser reescritos como

Passo E

Usar os pontos de quadratura $\bar{\boldsymbol{\theta}}_l = (\bar{\theta}_{l1}, \dots, \bar{\theta}_{lT})$, os pesos A_l , $l = 1, \dots, Q$, estimativas iniciais dos parâmetros dos itens, $\hat{\boldsymbol{\zeta}}_i$, $i = 1, \dots, n$, e dos parâmetros populacionais, $\hat{\boldsymbol{\eta}}$, para gerar $g_j^*(\bar{\boldsymbol{\theta}}_l)$ e, posteriormente, $r_{lit} = r_{it}(\boldsymbol{\theta}_l)$ e $f_l = f(\boldsymbol{\theta}_l)$, $i = 1, \dots, n$.

Passo M

Com $\mathbf{r} = (r_{lit})$, $\mathbf{f} = (f_{lit})$ obtidos nos Passos E, resolver as equações de estimação para $\hat{\boldsymbol{\eta}}$ e $\hat{\boldsymbol{\zeta}}_i$, $i = 1, \dots, n$, usando o algoritmo Newton-Raphson ou “Scoring” de Fisher (ver Apêndice A.3).

Estes passos compõem cada iteração do algoritmo EM, as quais serão repetidas até que algum critério de parada seja alcançado. Na situação em que precisamos estimar apenas os parâmetros dos itens, este algoritmo também pode ser aplicado, bastando desconsiderar a estimação de $\boldsymbol{\eta}$.

Com o processo de estimação dos parâmetros dos itens e populacionais concluído, restam alguns comentários. Para alguns conjuntos de dados pode haver uma certa instabilidade no

processo de estimação, levando alguns itens a terem seus parâmetros muito mal estimados. Exemplos disso são $a_i < 0$ ou $c_i \notin [0, 1]$. Uma forma de contornar essa situação é utilizar uma abordagem bayesiana que atribua distribuições *a priori* aos parâmetros, outra é utilizar restrições nos parâmetros.

Um outro ponto de interesse é a estimação das habilidades individuais. Em algumas aplicações não há interesse na estimação dessas habilidades; apenas os parâmetros populacionais são de interesse. Em outras, os indivíduos podem fazer parte de algum processo seletivo, de forma que suas habilidades devem ser estimadas. A estimação bayesiana dos parâmetros dos itens, bem como a estimação das habilidades, formam o conteúdo do próximo capítulo.

Modelo Longitudinal para Uma Única População: Estimação Bayesiana - Parâmetros dos Itens e Populacionais e Habilidades

Neste capítulo tratamos, em uma primeira etapa, da estimação bayesiana dos parâmetros dos itens e parâmetros populacionais. Após esta etapa, pode haver interesse na estimação das habilidades, e essa estimação pode ser feita por máxima verossimilhança (ver, por exemplo, Andrade, Tavares & Valle, 2000) ou pela distribuição da habilidade, condicional aos vetores de observações individuais. Este último método, por usar o teorema de Bayes, normalmente é referido por *método bayesiano* para estimação das habilidades. Por conta disso, esta etapa de estimação também comporá este capítulo. A Seção 5.1 trata da estimação dos parâmetros dos itens e populacionais, enquanto a Seção 5.2 trata da estimação das habilidades.

5.1 Estimação dos Parâmetros dos Itens e Populacionais

O método de máxima verossimilhança apresenta problemas na estimação de itens que são respondidos corretamente, ou incorretamente, por todos os indivíduos, e também das habilidades de indivíduos que responderam corretamente, ou incorretamente, a todos os itens (ver Andrade, Tavares & Valle, 2000). Além disso, há a possibilidade de que as estimativas dos parâmetros dos itens caiam fora do intervalo esperado, tal como valores de a_i negativos, ou valores de c_i fora do intervalo $[0, 1]$. A metodologia bayesiana apresenta uma solução em que estes problemas são contornados.

Há várias propostas para a estimação bayesiana dos parâmetros de interesse na TRI. A mais utilizada é a *Estimação Bayesiana Marginal* proposta por Mislevy (1986), que é uma generalização da estimação por máxima verossimilhança marginal aplicada nos Capítulos 2 e 3. Basicamente, a estimação bayesiana consiste em estabelecer distribuições *a priori* para os parâmetros de interesse, construir uma nova função denominada distribuição *a posteriori* e estimar os parâmetros com base em alguma característica dessa distribuição. Consideremos que as componentes de $(\zeta', \eta)'$ são variáveis aleatórias independentes e contínuas, com distribuições especificadas. Por estarmos tratando de uma extensão da

estimação por máxima verossimilhança marginal, a estimação bayesiana também consiste de duas etapas.

Uma diferença que deve ser notada nesta seção, é que trataremos conjuntamente da obtenção dos parâmetros dos itens e populacionais, ao invés de compor blocos separados, como vinha acontecendo nos capítulos anteriores. Entretanto, no conhecimento dos parâmetros dos itens as equações de estimação dos parâmetros populacionais devem ser ignoradas, e vice-versa.

Consideremos que a distribuição de ζ_i , $i = 1, \dots, n$, é função de um vetor de parâmetros τ , com densidade $f(\zeta_i|\tau)$, e que a distribuição de η é função de um vetor de parâmetros φ , com densidade $h(\eta|\varphi)$. Com isso, a densidade conjunta desses parâmetros é

$$f(\theta, \zeta, \tau, \eta, \varphi) = f(\zeta|\tau)g(\theta|\eta)h(\eta|\varphi).$$

Para fazer inferências com relação a ζ e η , devemos nos basear na distribuição *a posteriori* de ζ e η , dada por

$$\begin{aligned} f^*(\zeta, \eta|U_{\dots}) &= C \int_{\mathbb{R}^T} P(U_{\dots}|\theta, \zeta) f(\zeta|\tau) g(\theta|\eta) h(\eta|\varphi) d\theta \\ &= C f(\zeta|\tau) h(\eta|\varphi) \left\{ \int P(U_{\dots}; \theta, \zeta) g(\theta|\eta) d\theta \right\} \\ &\propto L(\zeta, \eta) f(\zeta|\tau) h(\eta|\varphi), \end{aligned} \quad (5.1)$$

onde C representa uma constante e $L(\zeta, \eta) \equiv P(U_{\dots}|\zeta, \eta)$.

Como estimador de ζ podemos escolher alguma característica de $f^*(\zeta, \eta|U_{\dots})$, sendo que as mais adotadas são a média e a moda. No que segue vamos considerar a *moda da posteriori* como o estimador de ζ , ou seja, o valor de ζ que maximiza a posteriori marginal. Veremos que esta escolha tem a vantagem de incorporar expressões já obtidas anteriormente. Temos que

$$\log f^*(\zeta, \eta|U_{\dots}) = Const + \log L(\zeta, \eta) + \log f(\zeta|\tau) + \log h(\eta|\varphi),$$

onde o primeiro termo representa uma constante. Notando que a última parcela não é função de ζ_i , temos que as equações de estimação para os parâmetros dos itens ζ_i , $i = 1, \dots, n$, são dadas por

$$\frac{\partial f^*(\zeta, \eta|U_{\dots})}{\partial \zeta_i} = \frac{\partial \log L(\zeta, \eta)}{\partial \zeta_i} + \frac{\partial \log f(\zeta|\tau)}{\partial \zeta_i} = \mathbf{0}. \quad (5.2)$$

De forma similar, as equações de estimação para os parâmetros populacionais são dadas por

$$\frac{\partial f^*(\zeta, \eta | U_{..})}{\partial \eta} = \frac{\partial \log L(\zeta, \eta)}{\partial \eta} + \frac{\partial \log h(\eta | \varphi)}{\partial \eta} = \mathbf{0}. \quad (5.3)$$

A primeira parcela de (5.2) é exatamente a mesma obtida em (2.20). A abordagem bayesiana adiciona uma nova parcela a (2.20) relativa à distribuição *a priori* associada aos parâmetros dos itens. A primeira parcela de (5.2) relativa às componentes de ζ_i para o ML3 é dada por (2.24) a (2.26). A segunda parcela de (5.2) depende da priori adotada para cada parâmetro. Como espera-se que o parâmetro a_i seja positivo, b_i pode assumir qualquer valor real e c_i deve estar no intervalo $[0, 1]$, deveremos assumir distribuições que levam em conta essas limitações e isso exige um tratamento diferenciado para cada um destes parâmetros. Em seguida trataremos destes casos, considerando as suposições mais freqüentes na prática.

Distribuição a priori para os parâmetros dos itens

Priori para a_i

Geralmente, adotam-se as distribuições *Log-normal* ou *Qui-Quadrado* para a_i . Neste texto, vamos supor que cada parâmetro a_i tem distribuição Log-normal com parâmetro $\tau_a = (\mu_a, \sigma_a^2)$. Uma justificativa teórica para a adoção desta distribuição é que na prática os a_i são, em geral, positivos, sugerindo que a distribuição de a_i pode ser modelada por uma distribuição unimodal e com assimetria positiva (ver Mislevy (1986)), tal como a log-normal. A transformação $\alpha_i = \log a_i$ implica que cada α_i tem uma distribuição Normal $(\mu_\alpha, \sigma_\alpha^2)$, onde $\mu_a = \exp[\mu_\alpha + \sigma_\alpha^2/2]$ e $\sigma_a^2 = (\exp(\sigma_\alpha^2) - 1) \exp[2\mu_\alpha + \sigma_\alpha^2]$. Alguns autores (ver Baker (1992), por exemplo) preferem desenvolver expressões para a estimação de α_i ao invés de a_i e sugerem a utilização da propriedade de invariância do estimador de máxima verossimilhança para a obtenção de \hat{a}_i pela transformação $\hat{a}_i = \exp(\hat{\alpha}_i)$. Entretanto, para uniformidade desse texto, vamos continuar apresentando a equação para o parâmetro a_i .

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

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

Segue que a segunda parcela de (5.2), relativa ao parâmetro a_i , pode ser escrita como

$$\frac{\partial \log f(a_i | \mu_a, \sigma_a^2)}{\partial a_i} = -\frac{1}{a_i} \left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2} \right]. \quad (5.4)$$

Priori para b_i

Como os parâmetros de dificuldade estão na mesma escala da habilidade, em geral supõem-se que cada b_i tem distribuição *Normal* com vetor de parâmetros $\tau_b = (\mu_b, \sigma_b^2)'$. Desta forma, a segunda parcela de (5.2) pode ser escrita como

$$\frac{\partial \log f(b_i | \mu_b, \sigma_b^2)}{\partial b_i} = -\frac{(b_i - \mu_b)}{\sigma_b^2}. \quad (5.5)$$

Priori para c_i

Como c_i só pode pertencer ao intervalo $[0; 1]$, uma priori *Beta* foi proposta por Swaminathan & Gifford (1983). A função densidade da distribuição Beta com parâmetros $s + 1$ e $t + 1$ é dada por

$$f(c_i | s, t) = \frac{\Gamma(s + t + 2)}{\Gamma(s + 1)\Gamma(t + 1)} c_i^s (1 - c_i)^t, \quad (5.6)$$

onde $\Gamma(d)$ é a função Gama. A média desta distribuição é dada por

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

Swaminathan & Gifford propõem, ainda, a seguinte reparametrização:

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

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

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

Retornando a (5.6), obtemos

$$f(c_i | \alpha, \beta) = \frac{\Gamma(\alpha + \beta - 2)}{\Gamma(\alpha - 1)\Gamma(\beta - 1)} c_i^{\alpha-2} (1 - c_i)^{\beta-2}. \quad (5.7)$$

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

Para chegarmos à expressão para a segunda parcela de (5.2), notemos que

$$\log f(c_i|\alpha, \beta) = Const + (\alpha - 2) \log c_i + (\beta - 2) \log(1 - c_i). \quad (5.8)$$

Conseqüentemente,

$$\frac{\partial \log f(c_i|\alpha, \beta)}{\partial c_i} = \frac{\alpha - 2}{c_i} - \frac{\beta - 2}{1 - c_i}. \quad (5.9)$$

Com as componentes (5.4), (5.5) e (5.9), temos que as equações de estimação para as componentes de ζ_i são

$$a_i : D(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it})(\theta_t - b_i) W_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} - \frac{1}{a_i} \left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2} \right] = 0, \quad (5.10)$$

$$b_i : -Da_i(1 - c_i) \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} - \frac{(b_i - \mu_b)}{\sigma_b^2} = 0, \quad (5.11)$$

$$c_i : \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^*} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} + \frac{\alpha - 2}{c_i} - \frac{\beta - 2}{1 - c_i} = 0. \quad (5.12)$$

Para efeito de aplicação dos procedimentos iterativos Newton-Raphson ou “Scoring” de Fisher, precisaremos das derivadas segundas das expressões (5.10) a (5.12). Como as derivadas segundas das primeiras parcelas dessas expressões já foram obtidas no Capítulo 2, resta apenas a obtenção para as segundas parcelas, que são as seguintes:

$$\begin{aligned} \frac{\partial^2 \log f(a_i|\mu_a, \sigma_a^2)}{\partial a_i^2} &= \frac{1}{a_i \sigma_a^2} [\sigma_a^2 + \log a_i - \mu_a - 1], \\ \frac{\partial^2 \log f(b_i|\mu_b, \sigma_b^2)}{\partial b_i^2} &= -\frac{1}{\sigma_b^2}, \\ \frac{\partial^2 \log f(c_i|\alpha, \beta)}{\partial c_i^2} &= -\frac{\alpha - 2}{c_i^2} - \frac{\beta - 2}{(1 - c_i)^2}. \end{aligned} \quad (5.13)$$

Distribuição a priori para os parâmetros populacionais

Embora a estimação dos parâmetros populacionais não padeça da instabilidade encontrada na estimação dos parâmetros dos itens por máxima verossimilhança, algum conhecimento a priori pode vir a ser disponível para estimação dos parâmetros populacionais e

isso será considerado nesta sessão. Além disso, em situações em que precisamos estimar algum parâmetro cujo espaço paramétrico é um subconjunto de \mathbb{R} , tal como $\rho \in [-1, 1]$, pode vir a ser muito útil encontrar meios de manter as estimativas no intervalo desejado.

Pela expressão (5.3), precisamos apenas $\partial \log h(\boldsymbol{\eta}|\boldsymbol{\varphi})/\partial \boldsymbol{\eta}$, onde $\boldsymbol{\varphi}$ é conhecido. Porém, a suposição de prioris depende de qual conjunto de parâmetros $\boldsymbol{\eta}$ estejamos adotando. Para exemplificar, consideremos que $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Para o vetor de médias podemos propor

$$\boldsymbol{\mu} \sim N_T(\boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_\mu).$$

As prioris para a matriz de covariância $\boldsymbol{\Sigma}$ dependem de qual estrutura estejamos adotando para ela. Para as matrizes Uniforme, Bandas e AR(1), com $\sigma^2 = 1$, só temos o parâmetro ρ , que em muitos casos está no intervalo $[0, 1]$. Nesta situação pode-se propor uma priori $Beta(\alpha, \beta)$ para ρ . A *fdp* da $Beta(\alpha, \beta)$ escrita como $f((y+1)/2 \equiv x|\alpha, \beta)$ tem suporte no intervalo $y \in [-1, 1]$, de forma que ela pode ser usada quando ρ puder ser negativo, mas alternativas mais adequadas podem ser encontradas em Bernardo & Smith (1994). As matrizes Diagonal e Hankel têm parâmetros, σ_t^2 , $t = 1, \dots, T$, que devem ser positivos e por isso podem ter associadas prioris *Inversa-Gama* ou *Qui-Quadrado*. A priori para o parâmetro σ_{12} da matriz de Hankel deve levar em conta o fato de que ele é limitado, pois $\sigma_{12} \in [-m, m]$, onde $m = \min_{t,s} \sqrt{\sigma_{tt}\sigma_{ss}}$.

5.2 Estimação das Habilidades

Nesta seção tratamos da estimação dos vetores de habilidades individuais $\boldsymbol{\theta}_j = (\theta_{j1}, \dots, \theta_{jT})'$. Como comentado na Seção 1.2, a estimação por máxima verossimilhança marginal consiste de duas etapas: primeiro a estimação dos parâmetros dos itens e populacionais, e posteriormente, das habilidades. Na estimação por máxima verossimilhança marginal supomos que as habilidades da população nas T condições de avaliação seguem uma distribuição cuja *fdp* é $g(\boldsymbol{\theta}|\boldsymbol{\eta})$, a partir da qual podemos estimar $\boldsymbol{\eta}$, o que não poderia ser feito na estimação em uma etapa porque $\boldsymbol{\eta}$ é o conjunto de parâmetros da função g .

Como nos Capítulos 2 a 4 já tratamos da estimação dos parâmetros dos itens e parâmetros populacionais, estes serão considerados conhecidos.

A estimação da habilidade de um indivíduo pode ser tratada de várias formas; as principais são por *máxima verossimilhança* e *bayesiana*, esta última através da distribuição da habilidade, dado o conjunto de observações do referido indivíduo. A estimação por máxima verossimilhança consiste em trabalhar com a expressão (2.4), referida por *verossimilhança individual*. Entretanto, a habilidade de um indivíduo no teste t pode ser melhor estimada se houver alguma informação oriunda das habilidades nos outros testes. Essa informação será dada pela estrutura de correlação entre as habilidades nos T testes. Por conta disso, trataremos da estimação conjunta, ou seja, do vetor das habilidades nos T testes.

Para incorporar a informação contida na estrutura de covariâncias entre as habilidades nos T testes, a estimação de $\boldsymbol{\theta}_j$ será feita com base na distribuição condicional de $\boldsymbol{\theta}_j$,

dados o conjunto de respostas deste indivíduo, os parâmetros dos itens e populacionais, ou seja,

$$g_j^*(\boldsymbol{\theta}_j) = \mathbb{P}(\boldsymbol{\theta}_j | \mathbf{U}_{j..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \frac{P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta}) g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}, \quad (5.14)$$

que será tratada simplesmente como a *Condicional de $\boldsymbol{\theta}_j$* e que na abordagem bayesiana é a distribuição *a posteriori* de $\boldsymbol{\theta}_j$. As características de (5.14) mais utilizadas como estimadores de $\boldsymbol{\theta}_j$ são a moda e a média da condicional de $\boldsymbol{\theta}_j$. A seguir, trataremos da obtenção de cada uma destas características.

5.2.1 Moda da Distribuição Condicional de $\boldsymbol{\theta}_j$

O processo de estimação de $\boldsymbol{\theta}_j$ pela moda de sua condicional consiste em obter o máximo da função $g_j^*(\boldsymbol{\theta}_j)$. Temos que

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

Com isso, a equação de estimação será

$$\frac{\partial \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} = \frac{\partial \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \boldsymbol{\theta}_j} + \frac{\partial \log g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j} = \mathbf{0}. \quad (5.15)$$

Notemos que de (2.4),

$$\log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta}) = \sum_{t=1}^T \sum_{i \in \mathbf{I}_t} \log P(U_{jit} | \theta_{jt}, \zeta_i)$$

e

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

Mas

$$\begin{aligned}
\frac{\partial \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \theta_{jt}} &= \frac{\partial}{\partial \theta_{jt}} \sum_{i \in \mathbf{I}_t} \log P(U_{jit} | \theta_{jt}, \zeta_i) \\
&= \sum_{i \in \mathbf{I}_t} \frac{\partial P(U_{jit} | \theta_{jt}, \zeta_i) / \partial \theta_{jt}}{P(U_{jit} | \theta_{jt}, \zeta_i)} \\
&= \sum_{i \in \mathbf{I}_t} \frac{(-1)^{U_{jit}+1}}{P_{jit}^{U_{jit}} Q_{jit}^{1-U_{jit}}} \left(\frac{\partial P_{jit}}{\partial \theta_{jt}} \right) \\
&= \sum_{i \in \mathbf{I}_t} V_{jit} \left(\frac{\partial P_{jit}}{\partial \theta_{jt}} \right),
\end{aligned}$$

onde P_{jit} é dada por (2.1). Notemos que

$$\frac{\partial P_{jit}}{\partial \theta_{jt}} = -\frac{\partial P_{jit}}{\partial b_i} = Da_i(1 - c_i)P_{jit}^*Q_{jit}^* = \alpha_i P_{jit}^*Q_{jit}^*,$$

onde $\alpha_i = Da_i(1 - c_i)$ e P_{jit}^* é dada por (2.16) com θ_t substituído por θ_{jt} . Segue que

$$\frac{\partial \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \theta_{jt}} = \sum_{i \in \mathbf{I}_t} \alpha_i V_{jit} P_{jit}^* Q_{jit}^*. \quad (5.16)$$

Também,

$$\begin{aligned}
\frac{\partial \log g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j} &= -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}_j} (\boldsymbol{\theta}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_j - \boldsymbol{\mu}) \\
&= -\frac{1}{2} \times 2 \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_j - \boldsymbol{\mu}) \\
&= -\boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_j - \boldsymbol{\mu}).
\end{aligned} \quad (5.17)$$

Com as expressões (5.16) e (5.17) chegamos à equação de estimação para $\boldsymbol{\theta}_j$, $j = 1, \dots, N$. Infelizmente, a Equação (5.15) não possui solução explícita, logo precisaremos aplicar algum processo iterativo. A seguir, apresentaremos as expressões necessárias para aplicação do Procedimento Newton-Raphson.

Aplicação do Procedimento Newton-Raphson

Para aplicação do procedimento de Newton-Raphson será necessária a derivada segunda de $g_j^*(\boldsymbol{\theta}_j)$. Apresentaremos apenas as expressões principais; os detalhes podem ser encontrados no Apêndice A.4.

Temos que

$$\frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} = \frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} + \frac{\partial^2 \log g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j}, \quad (5.18)$$

onde

$$\frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \theta_{jt}^2} = \sum_{i \in \mathbf{I}_t} \alpha_i^2 V_{jit} [-V_{jit} P_{jit}^* Q_{jit}^* + (1 - 2P_{jit}^*)] P_{jit}^* Q_{jit}^*, \quad (5.19)$$

$$\frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \theta_{jt} \partial \theta_{js}} = 0, \quad \forall s \neq t.$$

e, de (5.17),

$$\frac{\partial^2 \log g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} = -\boldsymbol{\Sigma}^{-1}. \quad (5.20)$$

Deste modo, considerando

$$\mathbf{f}_\theta(\boldsymbol{\theta}_j) = \frac{\partial \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}, \quad \mathbf{H}_\theta(\boldsymbol{\theta}_j) = \frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j} \quad (5.21)$$

e $\widehat{\boldsymbol{\theta}}_j^{(k)}$ uma estimativa de $\boldsymbol{\theta}_j$ na iteraçã k , entã na iteraçã $k + 1$ teremos que

$$\widehat{\boldsymbol{\theta}}_j^{(k+1)} = \widehat{\boldsymbol{\theta}}_j^{(k)} - [\mathbf{H}_\theta(\widehat{\boldsymbol{\theta}}_j^{(k)})]^{-1} \mathbf{f}_\theta(\widehat{\boldsymbol{\theta}}_j^{(k)}).$$

5.2.2 Média da Distribuiçã Condicional de $\boldsymbol{\theta}_j$

Para um indivíduo selecionado aleatoriamente de uma populaçã com distribuiçã das habilidades $g(\boldsymbol{\theta} | \boldsymbol{\eta})$ e com vetor de respostas $\mathbf{U}_{j..}$, temos que o denominador de (5.14) é dado por

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

Assim, o estimador de $\boldsymbol{\theta}_j$, dado pela média da posteriori (freqüentemente referido por MAP - maximum a posteriori), será obtido por

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_j &= \mathbb{E}(\boldsymbol{\theta} | \mathbf{U}_{j..}, \boldsymbol{\zeta}, \boldsymbol{\eta}) \\
&= \int_{\mathbb{R}^T} \boldsymbol{\theta} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \frac{\int_{\mathbb{R}^T} \boldsymbol{\theta} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^T} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}}.
\end{aligned} \tag{5.22}$$

Esta forma de estimação tem a vantagem de ser calculada diretamente, não necessitando da aplicação de métodos iterativos. Por conta disso alguns autores (por exemplo, Mislevy & Stocking (1989)) recomendam esta escolha para a estimação das habilidades. Além disso, esta quantidade já apareceu no processo de estimação dos parâmetros populacionais, através da expressão (3.7), de forma que as habilidades individuais já podem estar disponíveis ao final da primeira etapa do processo de estimação .

Em tudo o que foi feito até agora, consideramos que a estrutura de dados é completa, ou seja, todos os indivíduos responderam a todos os itens. Entretanto, na prática não é esse o caso mais freqüente. No capítulo seguinte trataremos do caso em que temos observações incompletas.

Modelo Longitudinal para Uma Única População: Dados Incompletos

Neste capítulo faremos a extensão dos capítulos anteriores para o caso em que temos dados incompletos. Na prática, a presença de vetores de observações individuais incompletas é bastante freqüente. Isto pode ocorrer devido a vários fatores, como por exemplo:

- o item não foi alcançado;
- o item não foi apresentado ao indivíduo;
- o item foi apresentado ao indivíduo e ele não respondeu.

O primeiro caso ocorre quando o tempo destinado ao teste não é suficiente para que indivíduos com desempenho lento possam alcançar todos os itens. Alguns softwares, tal como o LOGIST (Wingersky, Barton & Lord, 1982), procuram detectar estes casos e passam a ignorar todos os itens após a última resposta encontrada. No que segue vamos considerar que o tempo de teste foi suficiente para que todos os indivíduos possam responder a todos os itens.

O segundo caso é conhecido como *Ausência de Resposta por Planejamento* e também ocorre com bastante freqüência na prática. Por exemplo, se quisermos utilizar $n = 200$ itens em cada um dos T testes, será impraticável apresentar esta quantidade de itens a cada indivíduo. Como alternativa, podemos utilizar um planejamento, como *Blocos Incompletos Balanceados - BIB* (ver Cochran & Cox, 1950), formando vários sub-testes, e apresentar uma quantidade praticável de itens a cada indivíduo. Desta forma, todos os itens serão apresentados, embora não conjuntamente.

Pelo que foi desenvolvido até agora, os itens comuns entre dois testes, que permitem que a equalização seja feita, devem ser respondidos pelos mesmos indivíduos, o que é bem aceito pela maioria dos profissionais das áreas de avaliação educacional, mas passível de críticas por outros. A abordagem que propomos, e que leva em conta a existência de dados incompletos, permite que seja utilizado um planejamento de forma que um mesmo item não seja respondido duas vezes pelo mesmo indivíduo, eliminando possíveis críticas ao procedimento. Por exemplo, um sub-grupo A_1 de indivíduos responde aos itens comuns no teste 1, enquanto os mesmos itens seriam respondidos no teste 2 por um sub-grupo A_2 formado por indivíduos não contidos em A_1 . Em particular, este é exatamente o caso da aplicação que faremos no Capítulo 9.

Para diferenciar quais itens foram apresentados ao indivíduo j no teste t usaremos a variável indicadora

$$X_{jit} = \begin{cases} 1, & \text{se o item } i \text{ foi apresentado ao indivíduo } j \text{ no teste } t; \\ 0, & \text{se o item } i \text{ não foi apresentado ao indivíduo } j \text{ no teste } t. \end{cases} \quad (6.1)$$

Com esta notação o desenvolvimento das equações de estimação torna-se relativamente simples. Algumas expressões definidas no Capítulo 2 serão adaptadas e as expressões finais para as equações de estimação para os parâmetros dos itens e populacionais serão ligeiramente alteradas.

Uma alteração que deve ser ressaltada na abordagem de dados incompletos se refere aos padrões de resposta. Agora temos três possíveis respostas para cada item: ‘correto’, ‘incorreto’ e ‘não apresentado’. Considerando um total de n_t itens no teste t e $n_c = \sum_{t=1}^T n_t$, teremos $S = 3^{n_c}$ possíveis padrões de resposta. Neste caso, a abordagem de padrões de resposta provavelmente não será vantajosa, de forma que devem ser utilizados os vetores de respostas individuais.

6.1 Estimação dos Parâmetros dos Itens

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

$$P(\mathbf{U}_{j,t}|\theta_t, \boldsymbol{\zeta}) = \prod_{i \in \mathbf{I}_t} P(U_{jit}|\theta_t, \boldsymbol{\zeta}_i)^{X_{jit}}. \quad (6.2)$$

Adicionalmente, com a independência temporal podemos escrever

$$P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) = \prod_{t=1}^T \prod_{i \in \mathbf{I}_t} P(U_{jit}|\theta_t, \boldsymbol{\zeta}_i)^{X_{jit}}. \quad (6.3)$$

A probabilidade marginal do vetor de respostas $\mathbf{U}_{j..}$ é dada por

$$P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) = \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta})g(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta}. \quad (6.4)$$

Embora esta expressão esteja mantendo a mesma notação de (2.5), precisamos ter em mente que (6.4) tem uma informação adicional, a presença ou ausência de respostas, que não se fez necessária em (2.5) e agora é fundamental.

Para a determinação das equações de estimação para os parâmetros dos itens, notemos de (2.10) que

$$\begin{aligned}
\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left\{ \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \right\} \\
&= \int_{\mathbb{R}^T} \left\{ \frac{\partial}{\partial \zeta_i} \sum_{t=1}^T \sum_{l \in \mathbf{I}_t} \log P(U_{jlt}|\theta_t, \zeta_l)^{X_{jlt}} \right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \\
&= \int_{\mathbb{R}^T} \left\{ \sum_{t=1}^T \sum_{l \in \mathbf{I}_t} X_{jlt} \frac{\partial}{\partial \zeta_i} \log P(U_{jlt}|\theta_t, \zeta_l) \right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta} \\
&= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{jit} \left(\frac{\partial P(U_{jit}|\theta_t, \zeta_i)/\partial \zeta_i}{P(U_{jit}|\theta_t, \zeta_i)} \right) \right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta}.
\end{aligned}$$

Usando as expressões (2.13) a (2.16) chegamos a

$$\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{jit} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta}. \quad (6.5)$$

E, finalmente,

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} = \sum_{j=1}^N \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{jit} (U_{jit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (6.6)$$

onde $g_j^*(\boldsymbol{\theta})$ é a função definida em (2.19) com $P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta})$ e $P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})$ substituídos por (6.3) e (6.4), respectivamente.

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

$$\begin{aligned}
a_i : \quad & D(1 - c_i) \sum_{j=1}^N \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{jit} (U_{jit} - P_{it}) (\theta_t - b_i) W_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, \\
b_i : \quad & -D a_i (1 - c_i) \sum_{j=1}^N \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{jit} (U_{jit} - P_{it}) W_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, \\
c_i : \quad & \sum_{j=1}^N \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{jit} (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^*} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0.
\end{aligned}$$

Novamente, precisamos aplicar algum método iterativo para obtenção das estimativas de máxima verossimilhança. Entretanto, as expressões para a derivada segunda da log-verossimilhança com relação a ζ_i e ζ_l são praticamente as mesmas apresentadas no Capítulo 2, com as expressões (6.2) a (6.4) no lugar das expressões (2.2) a (2.5). Por conta disso, elas não serão apresentadas aqui.

Um detalhe que merece ser ressaltado se refere à mudanças nas expressões necessárias para aplicação do algoritmo EM. As expressões em (4.5) passam a ser escritas como

$$f_{it}(\boldsymbol{\theta}) = \sum_{j=1}^N X_{jit} g_j^*(\boldsymbol{\theta}) \quad \text{e} \quad r_{it}(\boldsymbol{\theta}) = \sum_{j=1}^N X_{jit} U_{jit} g_j^*(\boldsymbol{\theta}), \quad (6.7)$$

que representam, respectivamente, o número de indivíduos em uma população de tamanho N que têm vetor de habilidades $\boldsymbol{\theta}$ respondendo ao item i do teste t , e o número de tais indivíduos que respondem *corretamente* ao item i no teste t . Os conjuntos $\mathbf{f} = (f_{it}(\boldsymbol{\theta}))$ e $\mathbf{r} = (r_{it}(\boldsymbol{\theta}))$ compõem as quantidades não-observadas do algoritmo EM.

6.2 Estimação dos Parâmetros Populacionais

A informação sobre quais itens foram apresentados a cada indivíduo foi incluída apenas na distribuição do vetor de respostas, $\mathbf{U}_{j,t}$, não tendo influência sobre a distribuição de $\boldsymbol{\theta}_j$. Por conta disso as expressões para a estimação dos parâmetros populacionais permanecem quase inalteradas. Como conclusão, temos que as equações de estimação para os parâmetros populacionais são dadas por (3.4), com as componentes originais substituídas por (6.2) a (6.4). Estas últimas expressões também devem ser levadas em consideração para a obtenção da derivada segunda da log-verossimilhança com relação a $\boldsymbol{\eta}$, a ser aplicada no processo iterativo. Esta equação é praticamente a mesma referente ao Capítulo 3, e por isso não será apresentada aqui.

Completamos aqui o processo de estimação para o caso em que temos uma única população submetida a T condições de avaliação. Entretanto, é freqüente o caso em que temos mais de uma população envolvida na análise, e por isso este caso merece ser explorado. Este será o nosso objetivo no próximo capítulo.

Modelo Longitudinal para Várias Populações

Neste capítulo tratamos do caso em que temos $K > 1$ populações bem definidas em estudo e em uma ordem em possamos rotulá-las por *população 1*, \dots , *população K*. Exemplos mais comuns dessa situação ocorrem quando as populações são referentes a diferentes graus de escolaridade, sexo, regiões, métodos de ensino etc. Nesta situação, cada grupo selecionado será submetido a todas as T condições de avaliação, o que não quer dizer que cada indivíduo será submetido a todas elas, pois poderemos estar na situação de dados incompletos por planejamento. O conjunto de itens aplicados a cada grupo pode não ser o mesmo, como ocorre quando as populações são definidas por diferentes graus de escolaridade. Entretanto, há necessidade de fazermos uma ligação entre todos os testes em todos os tempos e, por isso, em algum teste deveremos ter itens comuns aplicados às populações 1 e 2, por exemplo. Podemos, inclusive, ter um item aplicado a duas populações em testes diferentes. Novamente, vamos assumir uma certa ordenação no conjunto integral de itens de forma que possamos representá-los por ζ_i , $i = 1, \dots, n$. Ainda, vamos considerar apenas o caso da estimação conjunta, de forma que o estabelecimento da métrica será necessária. Serão tratados tanto o caso de dados completos quanto o de dados incompletos.

Para incluir o caso de dados incompletos, utilizaremos a mesma abordagem do Capítulo 6. Seja X_{kjit} a variável indicadora definida por

$$X_{kjit} = \begin{cases} 1, & \text{se o item } i \text{ foi apresentado ao indivíduo } j \text{ do grupo } k \text{ no teste } t; \\ 0, & \text{se o item } i \text{ não foi apresentado ao indivíduo } j \text{ do grupo } k \text{ no teste } t. \end{cases}$$

7.1 Modelo Probabilístico

Sejam U_{kjit} a resposta ao item i e θ_{kjt} a habilidade do indivíduo j do grupo k no teste t . Para efeito de aplicações usaremos o ML3 para a probabilidade de um indivíduo responder corretamente a um item, ou seja,

$$P(U_{kjit} = 1 | \theta_{kjt}, \zeta_i) = c_i + (1 - c_i) \{1 + e^{-Da_i(\theta_{kjt} - b_i)}\}^{-1}. \quad (7.1)$$

Todas as suposições associadas ao modelo para uma única população descritas na

Seção 2.1 continuam sendo requeridas no caso atual. Porém, a presença de várias populações nos leva à necessidade de mais uma suposição, felizmente com completa aceitação na prática. Estamos trabalhando com populações distintas e por isso suporemos que as respostas oriundas de indivíduos de grupos diferentes serão independentes.

7.2 Distribuição das Habilidades

A presença de populações independentes nos possibilita tratar cada população separadamente. Estaremos, assim, assumindo que as habilidades de populações distintas são não-correlacionadas.

Assumiremos que o vetor de habilidades da população k nos T testes, $\boldsymbol{\theta}_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kT})'$, tem distribuição contínua multivariada com vetor de parâmetros $\boldsymbol{\eta}_k$ de componentes finitas, e denotaremos sua $f dp$ por $g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)$.

No caso Normal T -variado temos que $\boldsymbol{\eta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, onde $\boldsymbol{\mu}_k = (\mu_{k1}, \mu_{k2}, \dots, \mu_{kT})'$ é o vetor de médias e $\boldsymbol{\Sigma}_k = (\sigma_{kts})_{t,s \leq T}$, é a matriz de covariâncias associada à população k . O conjunto integral de parâmetros populacionais pode ser representado por $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, onde

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1T} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{K1} & \mu_{K2} & \cdots & \mu_{KT} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_K \end{pmatrix}. \quad (7.2)$$

Temos agora um maior número de opções para o arbítrio da população referência, podendo ser qualquer uma das populações em qualquer um dos tempos. Por conveniência, vamos adotar o teste 1 da população 1 como referência, ou seja,

$$\mu_{11} = 0, \quad \sigma_{11}^2 = 1. \quad (7.3)$$

7.3 Construção da Verossimilhança

A presença de mais de uma população na análise nos leva a algumas alterações nas notações já definidas. Sejam N_k o número de indivíduos no grupo k e n_{kt} o número de itens aplicados ao grupo k no teste t , $k = 1, \dots, K$ e $t = 1, \dots, T$. Os vetores de respostas serão representados por

$\mathbf{U}_{kj,t} = (U_{kj1t}, \dots, U_{kj n_{kt}t})$ o vetor ($n_{kt} \times 1$) de respostas do indivíduo j do grupo k no teste t ,

$\mathbf{U}_{kj,\cdot} = (U_{kj,1}, \dots, U_{kj,T})$ o vetor ($n_{ck} \times 1$) de respostas do indivíduo j do grupo k em todos os testes, onde $n_{ck} = \sum_{t=1}^T n_{kt}$.

$\mathbf{U}_{k,\dots} = (U_{k1,\dots}, \dots, U_{k N_k,\dots})$ o vetor ($N_k n_{ck} \times 1$) de respostas do grupo k em todos os testes,

$\mathbf{U}_{\dots} = (\mathbf{U}_{1..}, \dots, \mathbf{U}_{K..})$ o vetor $(\sum_{k=1}^K N_k n_{ck} \times 1)$ total de respostas.

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

$$P(\mathbf{U}_{kj.t} | \theta_t, \boldsymbol{\zeta}) = \prod_{i \in \mathbf{I}_{kt}} P(U_{kjit} | \theta_t, \zeta_i)^{X_{kjit}}, \quad (7.4)$$

em que \mathbf{I}_{kt} , $t = 1, \dots, T$, $k = 1, \dots, K$ é o conjunto dos índices dos itens presentes no teste t para o grupo k . Adicionalmente, com a independência temporal podemos escrever

$$P(\mathbf{U}_{kj..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) = \prod_{t=1}^T \prod_{i \in \mathbf{I}_{kt}} P(U_{kjit} | \theta_t, \zeta_i)^{X_{kjit}}. \quad (7.5)$$

A probabilidade marginal do vetor de respostas $\mathbf{U}_{kj..}$ é dada por

$$P(\mathbf{U}_{kj..} | \boldsymbol{\zeta}, \boldsymbol{\eta}_k) = \int_{\mathbb{R}^T} P(\mathbf{U}_{kj..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}_k) d\boldsymbol{\theta}. \quad (7.6)$$

e a verossimilhança por

$$L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \prod_{k=1}^K \prod_{j=1}^{N_k} P(\mathbf{U}_{kj..} | \boldsymbol{\zeta}, \boldsymbol{\eta}_k).$$

Com essa expressão, estamos aptos para construir as equações de estimação para os parâmetros dos itens ζ_i , $i = 1, \dots, n$, e para os parâmetros populacionais $\boldsymbol{\eta}$. Este será o nosso objetivo nas próximas seções. Boa parte dos desenvolvimentos são muito similares aos dos capítulos anteriores, por isso apenas os resultados principais serão apresentados.

7.4 Estimação dos Parâmetros dos Itens

As equações de estimação para os parâmetros dos itens são dadas por (2.8), com

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left\{ \sum_{k=1}^K \sum_{j=1}^{N_k} \log P(\mathbf{U}_{kj..} | \boldsymbol{\zeta}, \boldsymbol{\eta}_k) \right\} \\ &= \sum_{k=1}^K \sum_{j=1}^{N_k} \frac{1}{P(\mathbf{U}_{kj..} | \boldsymbol{\zeta}, \boldsymbol{\eta}_k)} \frac{\partial P(\mathbf{U}_{kj..} | \boldsymbol{\zeta}, \boldsymbol{\eta}_k)}{\partial \zeta_i}. \end{aligned} \quad (7.7)$$

De forma similar ao desenvolvimento em (2.10), temos que

$$\frac{\partial P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)}{\partial \boldsymbol{\zeta}_i} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{kjit} \left(\frac{\partial P(U_{kjit}|\theta_t, \boldsymbol{\zeta}_i)/\partial \boldsymbol{\zeta}_i}{P(U_{kjit}|\theta_t, \boldsymbol{\zeta}_i)} \right) \right\} P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}_k) d\boldsymbol{\theta},$$

Utilizando as notações em (2.16) e (2.17) e os desenvolvimentos em (2.13) e (2.14), chegamos a

$$\frac{\partial P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)}{\partial \boldsymbol{\zeta}_i} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{kjit} (U_{kjit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}_k) d\boldsymbol{\theta}. \quad (7.8)$$

Considerando

$$g_{kj}^*(\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{\theta}|\mathbf{U}_{kj..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_k) = \frac{P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)}, \quad (7.9)$$

teremos que a equação de estimação (7.7) pode ser escrita como

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i} = \sum_{k=1}^K \sum_{j=1}^{N_k} \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{kjit} (U_{kjit} - P_{it}) \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right\} g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0. \quad (7.10)$$

7.4.1 Aplicação ao Modelo Logístico de 3 Parâmetros

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

$$\begin{aligned} a_i : \quad & D(1 - c_i) \sum_{k=1}^K \sum_{j=1}^{N_k} \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{kjit} (U_{kjit} - P_{it}) (\theta_t - b_i) W_{it} \right\} g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, \\ b_i : \quad & -D a_i (1 - c_i) \sum_{k=1}^K \sum_{j=1}^{N_k} \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{kjit} (U_{kjit} - P_{it}) W_{it} \right\} g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, \\ c_i : \quad & \sum_{k=1}^K \sum_{j=1}^{N_k} \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} X_{kjit} (U_{kjit} - P_{it}) \frac{W_{it}}{P_{it}^*} \right\} g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0. \end{aligned}$$

Estas equações não possuem solução explícita e por isso precisaremos de algum método iterativo para a obtenção das estimativas de máxima verossimilhança dos parâmetros dos itens. As expressões das derivadas segundas da log-verossimilhança na atual situação são bastante similares às apresentadas no Capítulo 4, e por isso não serão apresentadas aqui. Para detalhes, veja a Seção A.5.

7.5 Estimação dos Parâmetros Populacionais

A modelagem da distribuição das habilidades no caso de vários grupos tem alguns pontos que precisam ser bem explorados, pois podem levar a uma estimação dos parâmetros populacionais com nível bem maior de complexidade do que o caso de um único grupo. O conjunto total de parâmetros é dado por (7.2). Não há, inicialmente, o que explorar com relação à matriz de médias, mas há bastante no que se refere às matrizes de covariâncias. Através das restrições em (7.3) fixamos o teste 1 do grupo 1 como referência, ou seja, a restrição se deu parcialmente sobre a matriz Σ_1 . Entretanto, nenhuma restrição será necessária para qualquer outra matriz Σ_k , $k > 1$. Como conseqüência, as matrizes Σ_k , $k = 1, \dots, K$, poderão ser diferentes. Podemos até supor que as estruturas de covariâncias nos K grupos são de tipos diferentes, como os tipos definidos nas Seções 3.3 a 3.7, e nesse caso o número de parâmetros de Σ em (7.2) pode ser relativamente grande, causando uma maior exigência computacional. Um dos casos mais simples seria a escolha (suposição) de $\Sigma_k = \sigma_k^2 \mathbf{R}$, onde \mathbf{R} é, por exemplo, a matriz de correlação uniforme. Neste caso, cada população adicional teria apenas um novo parâmetro em Σ para ser estimado. Entretanto, essa suposição impõe uma mesma estrutura de correlação para todas as populações, o que pode não ser verdadeiro.

Devido às restrições em (7.3), as equações de estimação para os parâmetros da população referência são ligeiramente diferentes das demais. De forma geral, as equações de estimação para os parâmetros populacionais são

$$\frac{\partial \log L(\zeta, \eta)}{\partial \eta_k} = \mathbf{0}, \quad k = 1, \dots, K. \quad (7.11)$$

Para chegarmos à expressão da função de estimação em (7.11), notemos que

$$\frac{\partial \log L(\zeta, \eta)}{\partial \eta_k} = \sum_{j=1}^{N_k} \frac{1}{P(\mathbf{U}_{kj..} | \zeta, \eta_k)} \frac{\partial P(\mathbf{U}_{kj..} | \zeta, \eta_k)}{\partial \eta_k} \quad (7.12)$$

e

$$\begin{aligned} \frac{\partial P(\mathbf{U}_{kj..}|\zeta, \boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} &= \int_{\mathbb{R}^T} P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \zeta) \left(\frac{\partial g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}} \right) d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta}|\boldsymbol{\eta}_k) \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}} \right) d\boldsymbol{\theta}. \end{aligned} \quad (7.13)$$

Então, segue que a equação de estimação para $\boldsymbol{\eta}_k$ é

$$\begin{aligned} \frac{\partial \log L(\zeta, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}_k} &= \sum_{j=1}^{N_k} \frac{1}{P(\mathbf{U}_{kj..}|\zeta, \boldsymbol{\eta}_k)} \int_{\mathbb{R}^T} P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta}|\boldsymbol{\eta}_k) \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right) d\boldsymbol{\theta} \\ &= \sum_{j=1}^{N_k} \int_{\mathbb{R}^T} \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right) g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbf{0}. \end{aligned} \quad (7.14)$$

No caso em que $\boldsymbol{\theta}_k$ tem distribuição Normal T -variada com vetor de médias $\boldsymbol{\mu}_k$ e matriz de covariâncias $\boldsymbol{\Sigma}_k$, temos que

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

Considerando $\mathbf{V}_k = \boldsymbol{\Sigma}_k^{-1} = (v_{kij})$, temos que

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\mu}_{k(1)}} = \mathbf{V}_{k(1)} (\boldsymbol{\theta} - \boldsymbol{\mu}_k), \quad k = 1, \quad (7.15)$$

e

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\mu}_k} = \mathbf{V}_k (\boldsymbol{\theta} - \boldsymbol{\mu}_k), \quad k > 1, \quad (7.16)$$

onde $\boldsymbol{\mu}_{k(1)} = (\mu_{k2}, \mu_{k3}, \dots, \mu_{kT})'$ e $\mathbf{V}_{k(1)}$ é a matriz \mathbf{V}_k sem a primeira linha. Para fins de estimação de $\boldsymbol{\eta}_k$ ainda serão necessárias as equações relativas aos parâmetros da estrutura de covariância adotada. De forma geral, temos que

$$\frac{\partial}{\partial \sigma_{kij}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k) = -\frac{1}{2} \frac{\partial (\log |\boldsymbol{\Sigma}_k|)}{\partial \sigma_{kij}} - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}_k)' \left[\frac{\partial \mathbf{V}_k}{\partial \sigma_{kij}} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}_k). \quad (7.17)$$

Podemos, agora, usar as expressões desenvolvidas nas Seções 3.3 a 3.7 para obter as equações de estimação para cada uma das estruturas de covariâncias adotadas. Na Seção

A.2 temos as expressões para as derivadas segundas, relativas a cada estrutura de covariância, e na Seção A.5 temos as derivadas segundas da log-verossimilhança, necessárias ao processo de estimação.

A estimação bayesiana é feita como no Capítulo 5, com todas as prioris sugeridas lá. Apenas o número de parâmetros a estimar será maior.

Até aqui não houve nenhuma exploração relativa aos vetores de médias. No próximo capítulo voltaremos nossa atenção para possíveis estruturas de crescimento nos vetores de médias ao longo das T condições de avaliação.

Curvas de Crescimento

O objetivo deste capítulo se concentra na exploração de algumas estruturas para os parâmetros populacionais de locação. Estaremos assumindo que a distribuição das habilidades dos indivíduos do grupo k é Normal Multivariada com parâmetros $\boldsymbol{\eta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. O interesse agora se concentra no vetor de médias, e não na matriz de covariâncias. Como para cada grupo o vetor $\boldsymbol{\mu}_k = (\mu_{k1}, \mu_{k2}, \dots, \mu_{kT})'$ representa as habilidades médias ao longo das T condições de avaliação, na prática esperamos que estas médias sejam crescentes no tempo. Em algumas aplicações podemos até ter o caso em que tais habilidades médias sejam decrescentes ao longo do tempo, ou ainda crescente em um período e decrescente em outro. De acordo com isso podemos pensar em modelar a habilidade média ao longo do tempo através de algumas funções de interesse, tais como a função linear ou quadrática. Essa modelagem pode trazer uma redução substancial no número de parâmetros a estimar. Se tivermos $T > 2$ condições de avaliação e K grupos, teremos $T \times K - 1$ médias em (7.2) a serem estimadas devido à restrição em (7.3); se supormos que a variação das habilidades médias ao longo do tempo é linear teremos $2 \times K - 1$ parâmetros, que é menor do que o número de parâmetros inicial.

Vamos considerar que os testes 1 a T são realizados nos tempos t_1, t_2, \dots, t_T . Se estes testes forem realizados em intervalos de tempos regulares, podemos adotar $t_1 = 1, t_2 = 2, \dots, t_T = T$. De qualquer forma, podemos assumir que o teste 1 foi realizado no tempo 1, i.e., $t_1 = 1$, e os outros tempos dependerão dos intervalos em que tais testes foram realizados a partir de t_1 .

De forma geral, seja

$$\mu_{kt} = f_k(t|\boldsymbol{\eta}_{1k}), \quad (8.1)$$

onde f_k representa uma função contínua duplamente diferenciável e $\boldsymbol{\eta}_{1k}$ representando o vetor de p_k parâmetros da função. O vetor de médias nas T condições de avaliação pode, agora, ser escrito como

$$\boldsymbol{\mu}_k = (f_k(t_1), f_k(t_2), \dots, f_k(t_T))'.$$

Muitas funções têm sido propostas na teoria de curvas de crescimento em dados longitudinais (ver Lindsey, 1993, por exemplo). Na maioria das vezes algumas funções simples, tais como as funções linear ou quadrática são suficientes para modelar uma estrutura longitudinal, mas polinômios de ordem maior que dois e algumas funções não-lineares

têm sido aplicadas em várias situações. A Tabela 8.1 apresenta uma relação de funções amplamente utilizadas em várias situações.

Tabela 8.1 Modelos de curvas de crescimento para o vetor de médias da população k

Logístico Triplo	$\frac{\gamma_{1k}}{1+e^{-\alpha_{1k}(t-\beta_{1k})}} + \frac{\gamma_{2k}}{1+e^{-\alpha_{2k}(t-\beta_{2k})}} + \frac{\gamma_{3k}}{1+e^{-\alpha_{3k}(t-\beta_{3k})}}$
Gompertz	$\gamma_{1k} + \gamma_{2k}e^{-e^{-\alpha_k(t-\beta_k)}}$
Jenks	$\alpha_k + \beta_k t - e^{\gamma_k + \delta_k t}$
Count	$\alpha_k + \beta_k t + \gamma_k \log t$
Mitscherlich	$\alpha_k - \beta_k e^{-\gamma_k t}$

A partir do modelo Logístico Triplo obtêm-se outros dois modelos bastante utilizados: o modelo *Logístico Duplo* ($\gamma_{3k} = 0$) e o *Logístico* ($\gamma_{2k} = 0$, $\gamma_{3k} = 0$).

Uma formulação a partir do qual um número bastante grande de modelos de curva de crescimento podem ser obtidos foi proposta por Nelder (1961,1962), com base na equação diferencial

$$\frac{d\mu_{kt}}{dt} = \tau_{3k}\mu_{kt}[h(e^{\tau_{2k}}, \tau_{4k}) - h(\mu_{kt}, \tau_{4k})], \quad (8.2)$$

onde

$$h(\lambda, \tau_{4k}) = \begin{cases} \frac{\lambda^{\tau_{4k}} - 1}{\tau_{4k}} & \tau_{4k} \neq 0 \\ \log(\lambda) & \tau_{4k} = 0, \end{cases}$$

com condição inicial $\mu_{11} = \exp(\tau_{11})$ no ponto inicial $t = t_1 = 1$, cuja solução é dada por

$$\mu_{kt} = \begin{cases} e^{\tau_{2k}} [1 + (e^{(\tau_{2k}-\tau_{1k})\tau_{4k}} - 1)e^{-\tau_{3k}(t-t_1)e^{\tau_{2k}\tau_{4k}}}]^{-\frac{1}{\tau_{4k}}}, & \tau_{4k} \neq 0 \\ e^{\tau_{2k} + (\tau_{1k} - \tau_{2k})e^{-\tau_{3k}(t-t_1)}}, & \tau_{4k} = 0. \end{cases} \quad (8.3)$$

Temos, então, que $\tau_{11} = \log(\mu_{11})$ e quando $\tau_{3k} > 0$, o parâmetro $\tau_{2k} = \lim_{t \rightarrow \infty} \log(\mu_{kt})$ representa a assíntota. Os parâmetros τ_{3k} e τ_{4k} controlam a taxa de crescimento de μ_{kt} . Se $\tau_{3k} > 0$ e $\tau_{2k} > \tau_{1k}$, teremos crescimento positivo de μ_{kt} , que é tipicamente o caso de maior interesse na maioria das situações práticas. Neste caso, a função assume o valor $e^{\tau_{1k}}$ em $t = t_1$ e cresce para o valor $e^{\tau_{2k}}$ quando $t \rightarrow \infty$. Se tivermos $\tau_{3k} < 0$ e $\tau_{2k} > \tau_{1k}$ ou $\tau_{3k} > 0$ e $\tau_{2k} < \tau_{1k}$, teremos crescimento negativo (decaimento). O parâmetro τ_{4k} determina o tipo da curva, tendo como casos particulares as curvas de Mitscherlich ($\tau_{4k} = -1$), Gompertz ($\tau_{4k} = 0$), Logística ($\tau_{4k} = 1$) e a exponencial ($\tau_{4k} \rightarrow \infty$ e $h(e^{\tau_{2k}}, \tau_{4k}) \rightarrow \text{constante}$).

Para aplicarmos esta função ao presente trabalho será necessária uma pequena alteração. Podemos notar que a restrição $\mu_{11} = 0$ nos levaria a $\tau_{11} = -\infty$ e a expressão (8.3) seria

inútil. Para contornar esse problema podemos notar que $\mu_{1t} = e^{\tau_{11}}$ no ponto inicial $t_1 = 1$, logo basta que esse valor seja subtraído em (8.3) para chegarmos a $\mu_{11} = 0$, isto é,

$$\mu_{kt} = \begin{cases} e^{\tau_{2k}} [1 + (e^{(\tau_{2k} - \tau_{1k})\tau_{4k}} - 1)e^{-\tau_{3k}(t-t_1)e^{\tau_{2k}\tau_{4k}}}]^{-\frac{1}{\tau_{4k}}} - e^{\tau_{11}}, & \tau_{4k} \neq 0 \\ e^{\tau_{2k} + (\tau_{1k} - \tau_{2k})e^{\tau_{3k}(t-t_1)}} - e^{\tau_{11}}, & \tau_{4k} = 0. \end{cases} \quad (8.4)$$

Heitjan (1991a,1991b) argumenta que se alguma covariável, z_{kt} , for disponível, essa informação pode ser incorporada ao parâmetro de assíntota, τ_{2k} , em (8.3) supondo-se que este parâmetro é uma função do tempo, da seguinte forma:

$$\tau_{2kt} = \tau_{1k} + \log \left(\frac{2}{1 + e^{\beta_k z_{kt}}} \right). \quad (8.5)$$

Se $z_{kt} = 0$, a assíntota é uma constante na condição inicial; se $z_{kt} = z_k$ é constante no tempo, a média cresce ou decresce da sua condição inicial, $e^{\tau_{1k}}$, para $2e^{\tau_{1k}} / (1 + e^{\beta_k z_k})$. Se $\beta_k = 0$, nenhum crescimento ou decrescimento ocorre para o grupo k .

A seguir, tratamos genericamente da obtenção das equações de estimação para os parâmetros das curvas de crescimento. Após isso, consideramos os casos particulares das curvas de Heitjan e polinomial.

8.1 Equações de Estimação

Nosso objetivo agora será obter as equações de estimação para os p_k parâmetros associados a f_k , que serão representados por $\boldsymbol{\eta}_{1k} = (\eta_{1k1}, \eta_{1k2}, \dots, \eta_{1kp_k})'$. De forma geral, as equações de estimação para esse conjunto de parâmetros são dadas por (7.14), ou seja,

$$\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}_{1k}} = \sum_{j=1}^{N_k} \int_{\mathbb{R}^T} \left(\frac{\partial \log g(\boldsymbol{\theta} | \boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{1k}} \right) g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbf{0}, \quad (8.6)$$

onde $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, \boldsymbol{\eta}'_2, \dots, \boldsymbol{\eta}'_K)'$ com $\boldsymbol{\eta}_k = (\boldsymbol{\eta}'_{1k}, \boldsymbol{\eta}'_{2k})'$, no qual $\boldsymbol{\eta}_{2k}$ é o vetor de parâmetros associados à estrutura de covariância adotada. Definindo

$$\boldsymbol{\mu}_{ks}^{(1)} = \frac{\partial \boldsymbol{\mu}_k}{\partial \eta_{1ks}}, \quad \boldsymbol{\Delta}_k^{(1)} = \left(\boldsymbol{\mu}_{k1}^{(1)}, \boldsymbol{\mu}_{k2}^{(1)}, \dots, \boldsymbol{\mu}_{kp_k}^{(1)} \right),$$

$$m_{ks}^{(1)} = \boldsymbol{\mu}_{ks}^{(1)'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \quad \text{e} \quad \mathbf{m}_k^{(1)} = \left(m_{k1}^{(1)}, m_{k2}^{(1)}, \dots, m_{kp_k}^{(1)} \right)',$$

teremos que

$$\begin{aligned}
\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \eta_{1ks}} &= -\frac{1}{2} \frac{\partial}{\partial \eta_{1ks}} [(\boldsymbol{\theta} - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_k)] \\
&= -\frac{1}{2} \left[-2 \left(\frac{\partial \boldsymbol{\mu}_k}{\partial \eta_{1ks}} \right)' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\theta} + \frac{\partial (\boldsymbol{\mu}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k)}{\partial \eta_{1ks}} \right] \\
&= -\frac{1}{2} \left[-2 \boldsymbol{\mu}_{ks}^{(1)'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\theta} + 2 m_{ks}^{(1)} \right] \\
&= \boldsymbol{\mu}_{ks}^{(1)'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\theta} - m_{ks}^{(1)}.
\end{aligned}$$

Com isso,

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{1k}} = \boldsymbol{\Delta}_k^{(1)'} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\theta} - \mathbf{m}_k^{(1)}. \quad (8.7)$$

Com (8.7) aplicado à (8.6) chegamos à equação de estimação para $\boldsymbol{\eta}_{1k}$. Para efeito de implementação do método Newton-Raphson serão necessárias as derivadas segundas de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)$. Em seguida apresentamos as principais expressões, deixando os detalhes para a Seção A.6. Sejam

$$\begin{aligned}
\boldsymbol{\Delta}_{ks}^{(2)} &= \frac{\partial \boldsymbol{\Delta}_k^{(1)}}{\partial \eta_{1ks}}, \quad \boldsymbol{\Delta}_k^{(2)} = \begin{pmatrix} \boldsymbol{\Delta}_{k1}^{(2)} \\ \vdots \\ \boldsymbol{\Delta}_{kp_k}^{(2)} \end{pmatrix}, \\
\mathbf{m}_{ks}^{(2)} &= \frac{\partial \mathbf{m}_k^{(1)}}{\partial \eta_{1ks}} \quad \text{e} \quad \mathbf{M}_k^{(2)} = \frac{\partial \mathbf{m}_k^{(1)}}{\partial \boldsymbol{\eta}_{1k}} = \left(\mathbf{m}_{k1}^{(2)}, \mathbf{m}_{k2}^{(2)}, \dots, \mathbf{m}_{kp_k}^{(2)} \right).
\end{aligned}$$

Então,

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{1k} \partial \boldsymbol{\eta}_{1k}'} = (\boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1}) \otimes \boldsymbol{\Delta}_k^{(2)} - \mathbf{M}_k^{(2)},$$

onde \otimes representa o *Produto de Kronecker* definido por

$$(\boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1}) \otimes \boldsymbol{\Delta}_k^{(2)} = \begin{pmatrix} \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_{k1}^{(2)} \\ \vdots \\ \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_{kp_k}^{(2)} \end{pmatrix}.$$

Para a obtenção das derivadas segundas com relação a $\boldsymbol{\eta}_{1k}$ e $\boldsymbol{\eta}_{2k}$, onde $\boldsymbol{\eta}_{2k} = (\eta_{2k1}, \eta_{2k2}, \dots, \eta_{2kp_k})'$, sejam $\mathbf{V}_k = \boldsymbol{\Sigma}_k^{-1}$,

$$\mathbf{V}_{kr}^{(1)} = \frac{\partial \mathbf{V}_k}{\partial \eta_{2kr}}, \quad \mathbf{V}_k^{(1)} = \begin{pmatrix} \mathbf{V}_{k1}^{(1)} \\ \vdots \\ \mathbf{V}_{kq_k}^{(1)} \end{pmatrix},$$

e

$$\mathbf{m}_{2kr}^{(2)} = \frac{\partial \mathbf{m}_k^{(1)}}{\partial \eta_{2kr}} \quad \text{e} \quad \mathbf{M}_{2k}^{(2)} = \left(\mathbf{m}_{2k1}^{(2)}, \mathbf{m}_{2k2}^{(2)}, \dots, \mathbf{m}_{2kq_k}^{(2)} \right).$$

Então, segue de (8.7) que

$$\frac{\partial \log g(\boldsymbol{\theta} | \boldsymbol{\eta}_k)}{\partial \eta_{2kr} \partial \eta'_{1k}} = \boldsymbol{\theta}' \mathbf{V}_{kr}^{(1)} \boldsymbol{\Delta}_k^{(1)} - \mathbf{m}_{2kr}^{(2)'}$$

Com isso, temos que

$$\frac{\partial^2 \log g(\boldsymbol{\theta} | \boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{2k} \partial \boldsymbol{\eta}'_{1k}} = \boldsymbol{\theta}' \otimes \mathbf{V}_k^{(1)} \otimes \boldsymbol{\Delta}_k^{(1)} - \mathbf{M}_{2k}^{(2)'},$$

onde

$$\boldsymbol{\theta}' \otimes \mathbf{V}_k^{(1)} \otimes = \begin{pmatrix} \boldsymbol{\theta}' \mathbf{V}_{k1}^{(1)} \boldsymbol{\Delta}_k^{(1)} \\ \vdots \\ \boldsymbol{\theta}' \mathbf{V}_{kq_k}^{(1)} \boldsymbol{\Delta}_k^{(1)} \end{pmatrix}.$$

As expressões das derivadas segundas de $\log g(\boldsymbol{\theta} | \boldsymbol{\eta}_k)$ com relação a $\boldsymbol{\eta}_{2k}$ para as estruturas de covariâncias estudadas neste trabalho são dadas nas Seções 3.3 a 3.7.

Como comentado antes, algumas funções são de particular interesse nesse tipo de abordagem. Na próxima seção exploramos o caso em que f_k é a curva de Heitjan.

8.2 Curva de Crescimento de Heitjan

Neste seção consideramos que a função de crescimento das habilidades médias de cada grupo é dada pela curva de Heitjan, expressa por (8.4). Basicamente, por (8.7), o que precisamos são as derivadas de μ_{kt} com relação ao conjunto de parâmetros $\boldsymbol{\tau}_k = (\tau_{1k}, \tau_{2k}, \tau_{3k}, \tau_{4k})'$. Estas derivadas resultam em expressões um tanto longas, por isso adotaremos algumas notações para simplificá-las. Considerando

$$\begin{aligned}
h_{1kt} &= h_{1t}(\boldsymbol{\tau}_k) = e^{(\tau_{2k}-\tau_{1k})\tau_{4k}} - 1 \\
h_{2kt} &= h_{2t}(\boldsymbol{\tau}_k) = e^{-\tau_{3k}(t-t_1)e^{\tau_{2k}\tau_{4k}}} \\
h_{3kt} &= h_{3t}(\boldsymbol{\tau}_k) = e^{(\tau_{2k}-\tau_{1k})e^{-\tau_{3k}(t-t_1)}} \\
h_{kt} &= h_{1kt}h_{2kt},
\end{aligned}$$

podemos escrever (8.4) como

$$\mu_{kt} = \begin{cases} e^{\tau_{2k}} [1 + h_{kt}]^{-\frac{1}{\tau_{4k}}} - e^{\tau_{11}}, & \tau_{4k} \neq 0 \\ e^{\tau_{2k}} h_{3kt} - e^{\tau_{11}}, & \tau_{4k} = 0 \end{cases} \quad (8.8)$$

Sejam, ainda, para $i = 1, 2, 3, 4$,

$$\mu_{kt}^{(i)} = \frac{\partial \mu_{kt}}{\partial \tau_{ik}}, \quad h_*^{(i)} = \frac{\partial h_*}{\partial \tau_{ik}} \quad \text{e} \quad 1_{\{k=1\}} = \begin{cases} 1, & \text{se } k = 1, \\ 0, & \text{se } k \neq 1. \end{cases}$$

Com esta notação, as derivadas primeiras de μ_{kt} com relação a $\boldsymbol{\tau}_k$ são dadas por

$$\begin{aligned}
\mu_{kt}^{(1)} &= \begin{cases} -\tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1}-1} h_{kt}^{(1)} - e^{\tau_{11}} 1_{\{k=1\}}, & \tau_{4k} \neq 0 \\ e^{\tau_{2k}} h_{3kt}^{(1)} - e^{\tau_{11}} 1_{\{k=1\}}, & \tau_{4k} = 0 \end{cases} \\
\mu_{kt}^{(2)} &= \begin{cases} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1}} - \tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1}-1} h_{kt}^{(2)}, & \tau_{4k} \neq 0 \\ e^{\tau_{2k}} [h_{3kt} + h_{3kt}^{(2)}], & \tau_{4k} = 0 \end{cases} \\
\mu_{kt}^{(3)} &= \begin{cases} -\tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1}-1} h_{kt}^{(3)}, & \tau_{4k} \neq 0 \\ e^{\tau_{2k}} h_{3kt}^{(3)}, & \tau_{4k} = 0 \end{cases} \\
\mu_{kt}^{(4)} &= \tau_{4k}^{-2} e^{\tau_{2k}} (1 + h_{kt})^{-\tau_{4k}^{-1}} [\log(1 + h_{kt}) - \tau_{4k} (1 + h_{kt})^{-1} h_{3kt}^{(4)}]. \quad (8.10)
\end{aligned}$$

As expressões para as quantidades $h_*^{(i)}$ podem ser encontradas na Seção A.7. Com (8.9) a (8.10) podemos compor as equações de estimação para os parâmetros das curvas de crescimento. Para o processo de estimação, são necessárias as derivadas segundas de μ_{kt} com relação a $\boldsymbol{\tau}_k$. Essas derivadas também podem ser encontradas na Seção A.7.

Na seção seguinte exploramos o caso em que f_k é um polinômio de grau d_k (ou seja, $p_k = d_k + 1$) e obteremos as expressões da Seção 8.1 para este caso. Geralmente $d_k = 1$ e 2 são os casos de maior interesse.

8.3 Curva de Crescimento Polinomial

Nesta seção vamos considerar que a função de crescimento das habilidades médias de cada grupo ao longo do tempo pode ser modelada (ou aproximada) por uma função polinomial. Esse tipo de função muitas vezes pode ser usada quando o interesse se restringe ao intervalo de tempo envolvido no experimento, mas previsões para períodos fora desse intervalo podem não ser confiáveis. Sandland & McGilchrist (1979) discutem bastante esse tipo de modelo indicando pontos que podem torná-lo não atrativo.

Para o modelo polinomial, temos

$$\mu_{kt} \equiv f_k(t) = \alpha_{0k} + \alpha_{1k}(t-1) + \cdots + \alpha_{d_k k}(t-1)^{d_k} = \sum_{s=0}^{d_k} \alpha_{sk}(t-1)^s,$$

onde por (7.3) temos que $\alpha_{01} = 0$. O conjunto de parâmetros associados a f_k é $\boldsymbol{\eta}_{1k} = (\alpha_{0k}, \alpha_{1k}, \dots, \alpha_{d_k k})'$, i.e., $\eta_{1ks} = \alpha_{(s-1)k}$, e o vetor de médias nas T condições de avaliação pode ser escrito como

$$\begin{aligned} \boldsymbol{\mu}_k &= \begin{pmatrix} \alpha_{0k} + \alpha_{1k}(t_1 - 1) + \cdots + \alpha_{d_k k}(t_1 - 1)^{d_k} \\ \vdots \\ \alpha_{0k} + \alpha_{1k}(t_T - 1) + \cdots + \alpha_{d_k k}(t_T - 1)^{d_k} \end{pmatrix} \\ &= \alpha_{0k} \mathbf{l}_0 + \alpha_{1k} \mathbf{l}_1 + \cdots + \alpha_{d_k k} \mathbf{l}_{d_k} = \sum_{s=0}^{d_k} \alpha_{sk} \mathbf{l}_s, \end{aligned}$$

onde $\mathbf{l}_s = ((t_1 - 1)^s, (t_2 - 1)^s, \dots, (t_T - 1)^s)'$, com $s = 0, 1, \dots, d_k$ e assumindo que $0^0 = 1$. Usando a notação

$$\gamma_{ijk} = \mathbf{l}'_i \boldsymbol{\Sigma}_k^{-1} \mathbf{l}_j, \quad \boldsymbol{\gamma}_{sk} = (\gamma_{s0k}, \gamma_{s1k}, \dots, \gamma_{sd_k k})' \quad \text{e} \quad \boldsymbol{\Gamma}_k = (\gamma_{ijk}), \forall i, j,$$

temos que

$$\boldsymbol{\mu}_{ks}^{(1)} = \frac{\partial \boldsymbol{\mu}_k}{\partial \alpha_{sk}} = \mathbf{l}_s, \quad \boldsymbol{\Delta}_k^{(1)} = (\mathbf{l}_0, \mathbf{l}_1, \dots, \mathbf{l}_{d_k}) \equiv \mathbf{L}_k$$

$$\mathbf{m}_{ks}^{(1)} = \mathbf{l}'_s \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k = \sum_{r=0}^{d_k} \alpha_{rk} \mathbf{l}'_s \boldsymbol{\Sigma}_k^{-1} \mathbf{l}_r = \sum_{r=0}^{d_k} \alpha_{rk} \gamma_{srk} = \boldsymbol{\gamma}'_{sk} \boldsymbol{\eta}_{1k},$$

e portanto,

$$\mathbf{m}_k^{(1)} = \boldsymbol{\Gamma}_k \boldsymbol{\eta}_{1k}.$$

Podemos, então, escrever (8.7) como

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{1k}} = \mathbf{L}'_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\theta} - \boldsymbol{\Gamma}_k \boldsymbol{\eta}_{1k}. \quad (8.11)$$

A derivada segunda de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)$ é

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{1k} \partial \boldsymbol{\eta}'_{1k}} = -\frac{\partial}{\partial \boldsymbol{\eta}_{1k}} (\boldsymbol{\eta}'_{1k} \boldsymbol{\Gamma}_k) = -\boldsymbol{\Gamma}_k. \quad (8.12)$$

Temos ainda que

$$\mathbf{m}_{2kr}^{(2)} = \frac{\partial \mathbf{m}_k^{(1)}}{\partial \eta_{2kr}} = \left(\frac{\partial \boldsymbol{\Gamma}_k}{\partial \eta_{2kr}} \right) \boldsymbol{\eta}_{1k},$$

onde

$$\frac{\partial \gamma_{ijk}}{\partial \eta_{2kr}} = \mathbf{l}'_i \left(\frac{\partial \mathbf{V}_k^{(1)}}{\partial \eta_{2kr}} \right) \mathbf{l}_j. \quad (8.13)$$

Com estas expressões aplicadas à Equação (A.16) obtemos as estimativas de máxima verossimilhança dos novos parâmetros em $\boldsymbol{\eta}_{1k}$, bem como de $\boldsymbol{\eta}_{2k}$, para $k = 1, \dots, K$.

Concluimos aqui boa parte do trabalho. Tratamos, na presença de uma única população em estudo e com estrutura de dados completa, da estimação dos parâmetros dos itens e dos parâmetros populacionais, de forma separada e, posteriormente, conjunta. Abordamos ainda a estimação bayesiana e estimação das habilidades. Após essa etapa, relaxamos a exigência de dados completos. O passo seguinte foi considerar a presença de várias populações em estudo, e finalmente, a exploração de estruturas de crescimento para os vetores de médias. Resta, então, fazer uma “exploração numérica” dos resultados teóricos. No próximo capítulo nos dedicamos às aplicações, tanto a dados simulados como a dados reais.

Aplicações

Neste capítulo fazemos a aplicação dos modelos propostos em dados simulados e em dados reais. Esta etapa será dividida em vários passos; na Seção 9.1 consideramos apenas a estimação dos parâmetros populacionais, fixando os parâmetros dos itens; na Seção 9.2 tratamos da estimação dos parâmetros dos itens, fixando os parâmetros populacionais, e a estimação conjunta é abordada na Seção 9.3. Na Seção 9.4 fazemos uma aplicação a dados reais, obtidos com o projeto *Avaliação do processo de inovações no ciclo básico e seu impacto sobre a situação de ensino-aprendizagem na região metropolitana de São Paulo* (AVJU), da Secretaria de Estado da Educação de São Paulo, considerando a estimação conjunta e explorando as várias estruturas de covariâncias adotadas neste trabalho.

Para o estudo de simulação, consideramos a situação em que $N = 1000$ indivíduos são submetidos a testes em $T = 5$ condições de avaliação. Para a probabilidade de acerto a qualquer dos itens, foi utilizado o ML3. A distribuição $N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ será utilizada para o vetor de habilidades. O vetor de médias $\boldsymbol{\mu} = (0, 1, 2, 3, 4)'$ será adotado em todas as situações e os valores para os parâmetros da matriz de covariância dependerão da particular estrutura de covariância considerada. Na Tabela 9.1 temos a relação de todos os valores usados para os parâmetros populacionais. Vale notar que a estrutura de covariância de Bandas só é positiva definida quando $|\rho| \leq 0,5$, e por isso o valor máximo adotado para esse parâmetro será $\rho = 0,45$. Para a estrutura de Hankel deve-se ter $|\sigma_{12}| \leq \min_{t,s} \sqrt{\sigma_t^2 \sigma_s^2}$ e por isso o valor máximo adotado para esse parâmetro será $\sigma_{12} = 0,75$.

Tabela 9.1 *Valores dos parâmetros populacionais usados nas simulações*

	μ_1	μ_2	μ_3	μ_4	μ_5	ρ ou σ_{12}	σ^2 ou σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Diagonal	0	1	2	3	4		1	0,8	0,9	1,1	1,2
Uniforme	0	1	2	3	4	0,25 e 0,75	1				
Bandas	0	1	2	3	4	0,25 e 0,45	1				
AR(1)	0	1	2	3	4	0,25 e 0,75	1				
Hankel	0	1	2	3	4	0,25 e 0,75	1	0,8	0,9	1,1	1,2

Para manter um controle mais rígido sobre o processo de estimação, optamos por transformar os valores gerados para as habilidades de forma que os vetores $\boldsymbol{\theta} = (\theta_1, \dots, \theta_T)'$ tenham vetores de médias e matrizes de covariâncias o mais próximo possível dos val-

ores desejados. Para isso, após gerada a matriz $N \times T$ de habilidades, foram obtidos o vetor de médias, \mathbf{m} , e a matriz de covariâncias amostral, \mathbf{S} . A transformação $\boldsymbol{\theta}^* = \boldsymbol{\Sigma}^{1/2} \mathbf{S}^{-1/2} (\boldsymbol{\theta} - \mathbf{m}) + \boldsymbol{\mu}$ gera valores para as habilidades com os parâmetros desejados. Os resultados numéricos desse procedimento foram bastante satisfatórios.

Cada um dos 5 testes será constituído por 24 itens, com 6 itens comuns entre cada par de testes adjacentes, ou seja, o teste 1 será formado pelos itens 1 a 24, o teste 2 pelos itens 19 a 42, o teste 3 pelos itens 37 a 60, o teste 4 pelos itens 55 a 78 e, finalmente, o teste 5 pelos itens 73 a 96. Na Tabela 9.2 temos os valores dos parâmetros dos itens (na escala Normal, isto é $D = 1,7$) utilizados na simulação. Foram considerados itens com o parâmetro a_i variando de 0,6 (baixa discriminação) a 1,4 (alta discriminação), o parâmetro b_i variando de -0,7 a 4,7 e o parâmetro c_i assumindo apenas o valor 0,20. A alocação dos itens procurou evitar a aplicação de itens muito fáceis ou muito difíceis em cada um dos testes. Por exemplo, a aplicação de um item com parâmetro b_i igual a 4,7 no teste 1 poderia trazer problemas de estimação, pois espera-se que poucos indivíduos tenham habilidade em torno deste valor na condição de avaliação 1. O mesmo raciocínio se estende aos outros testes.

O software utilizado para a simulação foi o Ox (Doornik, 2000), no qual construiu-se um programa específico para a estimação dos parâmetros envolvidos nos modelos aqui propostos. Este programa está disponibilizado para possíveis interessados, juntamente com um sucinto manual, na home-page <http://www.ime.usp.br/~hert/tri>.

Como a estimação dos parâmetros populacionais compõe apenas um bloco para a estimação conjunta, optou-se por fixar os parâmetros relativos à condição de avaliação 1. Desta forma, em todas as situações estaremos adotando $\mu_1 = 0$ e $\sigma_1^2 = 1$.

9.1 Estimação dos Parâmetros Populacionais

Nesta seção apresentamos alguns resultados obtidos via simulação, considerando cada uma das estruturas de covariâncias adotadas. As Tabelas 9.3 e 9.4 apresentam os resultados (estimativas médias e desvios-padrão) obtidos em $M = 500$ iterações. Para cada estrutura de correlação, a linha superior da tabela equivale às estimativas médias, enquanto as linhas inferiores representam os desvios-padrão estimados.

Observando as Tabelas 9.3 e 9.4, vemos que as estimativas médias estão muito próximas dos verdadeiros valores, comprovando a eficácia do processo de estimação dos parâmetros populacionais. Tais estimativas médias se mostraram um pouco melhores nos componentes do vetor de médias $\boldsymbol{\mu}$ do que nos componentes das matrizes de covariâncias.

As estimativas médias para σ_t^2 no modelo Diagonal foram muito boas. Os resultados obtidos para ρ nos modelos Uniforme, Bandas e AR(1), apresentaram valores muito próximos dos verdadeiros quando $\rho = 0,25$. Para $\rho = 0,75$, tais valores foram ligeiramente inferiores. Com relação ao parâmetro σ_{12} da estrutura de Hankel, os resultados obtidos foram ligeiramente inferiores aos verdadeiros nos dois casos considerados. Ainda sob essa

Tabela 9.2 *Valores adotados para os parâmetros dos itens*

Item	Teste	a_i	b_i	c_i	Item	Teste	a_i	b_i	c_i
1	1	0,6	-0,7	0,2	51	3	1,4	2,1	0,2
2	1	1,0	-0,7	0,2	52	3	0,6	2,3	0,2
3	1	1,4	-0,7	0,2	53	3	0,6	2,5	0,2
4	1	0,6	-0,5	0,2	54	3	0,6	2,7	0,2
5	1	1,0	-0,5	0,2	55	3;4	1,0	2,3	0,2
6	1	1,4	-0,5	0,2	56	3;4	1,0	2,5	0,2
7	1	0,6	-0,3	0,2	57	3;4	1,0	2,7	0,2
8	1	1,0	-0,3	0,2	58	3;4	1,4	2,3	0,2
9	1	1,4	-0,3	0,2	59	3;4	1,4	2,5	0,2
10	1	0,6	-0,1	0,2	60	3;4	1,4	2,7	0,2
11	1	1,0	-0,1	0,2	61	4	0,6	2,9	0,2
12	1	1,4	-0,1	0,2	62	4	1,0	2,9	0,2
13	1	0,6	0,1	0,2	63	4	1,4	2,9	0,2
14	1	1,0	0,1	0,2	64	4	0,6	3,0	0,2
15	1	1,4	0,1	0,2	65	4	1,0	3,0	0,2
16	1	0,6	0,3	0,2	66	4	1,4	3,0	0,2
17	1	0,6	0,5	0,2	67	4	0,6	3,1	0,2
18	1	0,6	0,7	0,2	68	4	1,0	3,1	0,2
19	1;2	1,0	0,3	0,2	69	4	1,4	3,1	0,2
20	1;2	1,0	0,5	0,2	70	4	0,6	3,3	0,2
21	1;2	1,0	0,7	0,2	71	4	0,6	3,5	0,2
22	1;2	1,4	0,3	0,2	72	4	0,6	3,7	0,2
23	1;2	1,4	0,5	0,2	73	4;5	1,0	3,3	0,2
24	1;2	1,4	0,7	0,2	74	4;5	1,0	3,5	0,2
25	2	0,6	0,9	0,2	75	4;5	1,0	3,7	0,2
26	2	1,0	0,9	0,2	76	4;5	1,4	3,3	0,2
27	2	1,4	0,9	0,2	77	4;5	1,4	3,5	0,2
28	2	0,6	1,0	0,2	78	4;5	1,4	3,7	0,2
29	2	1,0	1,0	0,2	79	5	0,6	3,9	0,2
30	2	1,4	1,0	0,2	80	5	1,0	3,9	0,2
31	2	0,6	1,1	0,2	81	5	1,4	3,9	0,2
32	2	1,0	1,1	0,2	82	5	0,6	4,0	0,2
33	2	1,4	1,1	0,2	83	5	1,0	4,0	0,2
34	2	0,6	1,3	0,2	84	5	1,4	4,0	0,2
35	2	0,6	1,5	0,2	85	5	0,6	4,1	0,2
36	2	0,6	1,7	0,2	86	5	1,0	4,1	0,2
37	2;3	1,0	1,3	0,2	87	5	1,4	4,1	0,2
38	2;3	1,0	1,5	0,2	88	5	0,6	4,3	0,2
39	2;3	1,0	1,7	0,2	89	5	0,6	4,5	0,2
40	2;3	1,4	1,3	0,2	90	5	0,6	4,7	0,2
41	2;3	1,4	1,5	0,2	91	5	1,0	4,3	0,2
42	2;3	1,4	1,7	0,2	92	5	1,0	4,5	0,2
43	3	0,6	1,9	0,2	93	5	1,0	4,7	0,2
44	3	1,0	1,9	0,2	94	5	1,4	4,3	0,2
45	3	1,4	1,9	0,2	95	5	1,4	4,5	0,2
46	3	0,6	2,0	0,2	96	5	1,4	4,7	0,2
47	3	1,0	2,0	0,2					
48	3	1,4	2,0	0,2					
49	3	0,6	2,1	0,2					
50	3	1,0	2,1	0,2					

Tabela 9.3 *Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,25$*

	μ_2	μ_3	μ_4	μ_5	ρ ou σ_{12}	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Diagonal	1,0012 (0,0125)	2,0026 (0,0136)	3,0038 (0,0151)	4,0014 (0,0151)		0,8001 (0,0338)	0,9041 (0,0373)	1,0977 (0,0508)	1,1959 (0,0540)
Uniform	0,9986 (0,0140)	1,9984 (0,0153)	2,9983 (0,0147)	3,9986 (0,0144)	0,2583 (0,0094)				
Bands	0,9984 (0,0140)	1,9973 (0,0144)	2,9978 (0,0145)	3,9982 (0,0141)	0,2525 (0,0095)				
AR(1)	0,9984 (0,0140)	1,9977 (0,0144)	2,9979 (0,0146)	3,9983 (0,0140)	0,2544 (0,0103)				
Hankel	1,0034 (0,0126)	2,0025 (0,0145)	3,0020 (0,0154)	4,0012 (0,0158)	0,2287 (0,0093)	0,7218 (0,0306)	0,8173 (0,0347)	1,0046 (0,045)9	1,1011 (0,0512)

Tabela 9.4 *Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,75$*

	μ_2	μ_3	μ_4	μ_5	ρ ou σ_{12}	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Diagonal	1,0012 (0,0125)	2,0026 (0,0136)	3,0038 (0,0151)	4,0014 (0,0151)		0,8001 (0,0338)	0,9041 (0,0373)	1,0977 (0,0508)	1,1959 (0,0540)
Uniforme	0,9986 (0,0159)	1,9981 (0,0171)	2,9982 (0,0179)	3,9991 (0,0169)	0,7291 (0,0076)				
Bandas*	0,9992 (0,0141)	1,9975 (0,0147)	2,9976 (0,0138)	3,9985 (0,0139)	0,4430 (0,0072)				
AR(1)	0,9985 (0,0158)	1,9976 (0,0159)	2,9981 (0,0165)	3,9987 (0,0138)	0,7343 (0,0066)				
Hankel	0,9989 (0,0165)	1,9989 (0,0175)	2,9983 (0,0184)	3,9983 (0,0172)	0,7248 (0,0189)	0,8020 (0,0271)	0,8832 (0,0342)	1,0967 (0,0351)	1,1926 (0,0385)

* Neste caso foi usado $\rho = 0,45$.

estrutura, as componentes σ_t^2 apresentaram resultados muito bons quando $\sigma_{12} = 0,75$, com resultados um pouco inferiores no caso $\sigma_{12} = 0,25$.

No geral, a estimação apresentou resultados satisfatórios, mas ainda há o que melhorar, o que provavelmente será alcançado com melhorias na implementação computacional do processo de estimação.

9.2 Estimação dos Parâmetros dos Itens

Nesta seção tratamos da estimação dos parâmetros dos itens. Para a probabilidade de acerto a qualquer dos itens, foi utilizado o ML3. A estimação foi realizada através do algoritmo EM, como descrito no Capítulo 4.

A Tabela 9.5 apresenta as *somas dos quadrados dos desvios (SQD)* das estimativas

médias, obtidas em $M = 500$ iterações, com relação aos verdadeiros valores, associada a cada *tipo de parâmetro*: discriminação, dificuldade e acerto ao acaso. Para cada item e cada parâmetro deste, as M estimativas produzem uma estimativa da variância, que são somadas para obtenção da *variabilidade total* (VT) associada a cada tipo de parâmetro; esses resultados são apresentados na Tabela 9.6. Tabelas contendo tais estimativas médias, bem como os desvios-padrão estimados podem ser encontradas na homepage <http://www.ime.usp.br/~hert/tri>. Tais arquivos são apresentados em formato *xls* e *dvi*. Os valores adotados para os parâmetros populacionais são aqueles apresentados na Tabela 9.1. Para o modelo de Hankel, o valor adotado para σ_{12} será o mesmo adotado para ρ nos outros modelos.

Tabela 9.5 *Soma dos quadrados dos desvios com relação aos verdadeiros valores.*

	$\rho = 0,25$			$\rho = 0,75$ (ou $\rho = 0,45$)		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Diagonal	0,1444	0,0761	0,0078			
Uniforme	0,3290	0,0561	0,0092	0,4483	0,0470	0,0121
Bandas*	0,3294	0,0561	0,0095	0,3429	0,0478	0,0089
AR(1)	0,3305	0,0583	0,0097	0,4216	0,0466	0,0106
Hankel	0,3542	0,0564	0,0096	0,5441	0,0520	0,0123

* A estrutura independe de ρ ** Neste caso foi usado $\rho = 0,45$ ao invés de $\rho = 0,75$ Tabela 9.6 *Soma das variâncias associadas a cada parâmetro*

	$\rho = 0,25$			$\rho = 0,75$ (ou $\rho = 0,45$)		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Diagonal*	2,8221	4,0138	0,5348			
Uniforme	1,9250	3,2079	0,4012	1,8898	3,4182	0,4246
Bandas**	1,9095	3,1904	0,3977	1,9050	3,2658	0,4074
AR(1)	1,8995	3,1709	0,3934	1,9288	3,3900	0,4205
Hankel	1,9238	3,2366	0,4062	1,8633	3,4773	0,4285

* A estrutura independe de ρ ** Neste caso foi usado $\rho = 0,45$ ao invés de $\rho = 0,75$

Podemos notar pela Tabela 9.5 que mudanças nas estruturas de covariância adotadas agem de forma diferente nos parâmetros dos itens. O parâmetro de discriminação foi melhor estimado quando $\rho = 0,25$. Entretanto, o oposto aconteceu com o parâmetro de dificuldade, que foi pior estimado quando $\rho = 0,25$. O parâmetro de acerto ao acaso foi melhor estimado quando $\rho = 0,25$, com exceção para a estrutura de Bandas. Pela Tabela 9.6, vemos que estas relações não se mantêm quando consideramos a variabilidade

total. A variabilidade associada ao parâmetro de discriminação foi maior quando $\rho = 0,25$, com exceção para o modelo AR(1). Para os parâmetros de dificuldade e acerto ao acaso, a variabilidade foi menor quando $\rho = 0,25$.

Enfim, quando o objetivo é acurácia na estimação, podemos dizer que correlações altas ajudam na estimação do parâmetro de dificuldade, mas prejudicam a estimação do parâmetro de discriminação. Essa conclusão inverte-se quando o objetivo é uma menor variabilidade. A estimação do parâmetro de acerto ao acaso é sempre favorecida por correlações baixas.

Vale a pena ressaltar que apesar das considerações feitas, tanto SQD quanto VT apresentam valores bem pequenos, mostrando a eficácia dos modelos propostos.

9.3 Estimação Conjunta

Nesta seção temos o caso mais freqüente que ocorre na prática, quando tem-se que estimar tanto os parâmetros dos itens quanto os populacionais. As Tabelas 9.7 e 9.8 apresentam as mesmas quantidades descritas nas seções anteriores, ou seja, as estimativas médias dos parâmetros populacionais e as somas dos quadrados dos desvios e variabilidade total, no caso dos parâmetros dos itens.

As estimativas das médias estão muito próximas dos verdadeiros valores, tanto para $\rho = 0,25$ quanto para $\rho = 0,75$. As estimativas médias para ρ e σ_{12} se mostraram melhores quando $\rho = 0,25$, com uma aparente subestimação no caso $\rho = 0,75$.

As quantidades σ_t^2 não foram muito bem estimadas para o modelo Diagonal. Para o modelo de Hankel, tais estimativas apresentaram-se muito próximas dos valores desejados quando $\sigma_{12} = 0,75$, mas não com $\sigma_{12} = 0,25$. Curiosamente, a estimação de σ_{12} teve resultados melhores do que os apresentados anteriormente.

Se compararmos os resultados obtidos nesta seção com aqueles da Seção 9.1, veremos que os vetores de médias foram bem estimados em ambos os casos, e as quantidades ρ e σ_{12} apresentam conclusões parecidas nas duas seções. Entretanto, foram detectados vícios associados a σ_t^2 no modelo Diagonal que antes não haviam aparecido. No modelo de Hankel também apareceu um vício quando $\rho = 0,25$, equivalente ao dobro (fator 2: relação entre duas quantidades) do apresentado anteriormente. Ainda, os desvios-padrão associados aos parâmetros populacionais mostraram-se bem maiores no caso da estimação conjunta, com fatores superiores a 5, relativos às habilidades médias no modelo Diagonal, mas ficando em torno de 3 nos outros casos. Novamente, ainda há bastante a melhorar, o que provavelmente será alcançado com mais investimento na implementação computacional do processo de estimação.

Tabela 9.7 *Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,25$*

	μ_2	μ_3	μ_4	μ_5	ρ ou σ_{12}	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Diagonal	0,99345 (0,0385)	1,9957 (0,0579)	3,0123 (0,0725)	4,0350 (0,0862)		0,7961 (0,0697)	0,9113 (0,0822)	1,1446 (0,0854)	1,2580 (0,1380)
Uniforme	1,0168 (0,0344)	2,0145 (0,0471)	3,0086 (0,0576)	4,0037 (0,0664)	0,2569 (0,0093)				
Bandas	1,0154 (0,0352)	2,0145 (0,0485)	3,0099 (0,0576)	4,0060 (0,0682)	0,2526 (0,0101)				
AR(1)	1,0154 (0,0349)	2,0142 (0,0482)	3,0093 (0,0581)	4,0050 (0,0693)	0,2542 (0,0103)				
Hankel	1,0394 (0,0345)	2,0863 (0,0481)	3,0939 (0,0585)	4,0544 (0,0662)	0,2417 (0,0127)	0,7947 (0,0488)	0,8612 (0,0488)	0,9653 (0,0501)	1,0125 (0,0586)

Tabela 9.8 *Estimativas médias (desvios-padrão): $\rho = \sigma_{12} = 0,75$*

	μ_2	μ_3	μ_4	μ_5	ρ ou σ_{12}	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Diagonal	0,99345 (0,0385)	1,9957 (0,0579)	3,0123 (0,0725)	4,0350 (0,0862)		0,7961 (0,0697)	0,9113 (0,0822)	1,1446 (0,0854)	1,2580 (0,1380)
Uniforme	1,0113 (0,0345)	2,0066 (0,0491)	3,0020 (0,0621)	3,9967 (0,0734)	0,7348 (0,0127)				
Bandas*	1,0158 (0,0345)	2,0149 (0,0484)	3,0074 (0,0587)	4,0065 (0,0703)	0,4429 (0,0082)				
AR(1)	1,0120 (0,0345)	2,0110 (0,0490)	3,0070 (0,0600)	4,0047 (0,0709)	0,7372 (0,0080)				
Hankel	1,0372 (0,0422)	2,0751 (0,0516)	3,0786 (0,0647)	4,0416 (0,0758)	0,7682 (0,0326)	0,8293 (0,0797)	0,9555 (0,0367)	1,1011 (0,0454)	1,1725 (0,0507)

* Neste caso foi usado $\rho = 0,45$.

Com relação aos parâmetros dos itens da Tabela 9.9, temos que o parâmetro de discriminação foi melhor estimado quando $\rho = 0,25$. O oposto aconteceu com o parâmetro de dificuldade, que foi pior estimado quando $\rho = 0,25$, com exceção do modelo de Bandas. O parâmetro de acerto ao acaso foi melhor estimado quando $\rho = 0,25$. Pela Tabela 9.10, vemos que estas relações não se mantêm quando consideramos a variabilidade total. A variabilidade associada ao parâmetro de discriminação foi maior quando $\rho = 0,25$, com exceção do modelo de Bandas. Para os parâmetros de dificuldade e acerto ao acaso a variabilidade foi menor quando $\rho = 0,25$.

Se compararmos os resultados obtidos nesta seção com aqueles da Seção 9.2, veremos que os resultados referentes às Tabelas 9.9 e 9.10 estão muito próximos na maioria dos casos dos resultados das Seções 9.5 e 9.6. No modelo de Hankel apenas o parâmetro de dificuldade foi afetado, com SQD equivalente a 5 vezes o apresentado anteriormente, mas sem

comprometimento na VT. O modelo de Bandas também foi bastante afetado, mas apenas quando $\rho = 0,45$, apresentando quantidades equivalentes ao triplo das apresentadas anteriormente. Com relação à VT, esse aumento ficou em torno de 30%.

Como esperado, os resultados obtidos com a estimação conjunta foram inferiores aos obtidos nas situações onde um grupo de parâmetros era conhecido. Entretanto, em ambos os casos as precisões foram boas.

Tabela 9.9 *Soma dos quadrados dos desvios com relação aos verdadeiros valores.*

	$\rho = 0,25$			$\rho = 0,75$		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Diagonal*	0,1930	0,1296	0,0162			
Uniforme	0,3288	0,0483	0,0093	0,4326	0,0447	0,0117
Bandas**	0,3322	0,0493	0,0098	0,9775	0,1761	0,0153
AR(1)	0,3289	0,0504	0,0099	0,4155	0,0414	0,0102
Hankel	0,3170	0,3320	0,0099	0,5982	0,2750	0,0113

* A estrutura independe de ρ

** Neste caso foi usado $\rho = 0,45$ ao invés de $\rho = 0,75$

Tabela 9.10 *Soma das variâncias associadas a cada parâmetro*

	$\rho = 0,25$			$\rho = 0,75$		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Diagonal	3,0975	4,2745	0,5333			
Uniforme	1,8798	3,2800	0,3862	1,8710	3,4937	0,4100
Bandas*	1,8724	3,2711	0,3835	2,6327	4,2738	0,4123
AR(1)	1,8796	3,2990	0,3858	1,8830	3,4605	0,4039
Hankel	1,9528	3,1561	0,3905	1,7978	3,4900	0,3957

* Neste caso foi usado $\rho = 0,45$ ao invés de $\rho = 0,75$

9.4 Aplicação a Dados Reais: Projeto AVJU

Nesta seção utilizaremos um conjunto de dados provenientes de um estudo longitudinal da Secretaria de Estado de Educação de São Paulo, denominado *Avaliação do processo de inovações no ciclo básico e seu impacto sobre a situação de ensino-aprendizagem na região metropolitana de São Paulo (AVJU)*, cujo objetivo foi acompanhar o desempenho de uma determinada população referente à habilidade em matemática. Um grupo de aproximadamente 3000 crianças foi submetido a $T = 3$ testes, realizados nos meses de novembro de 1992, 1993 e 1994. Na ocasião do teste 1, todos os indivíduos cursavam a primeira série do ensino fundamental.

O conjunto de dados é incompleto, no sentido de que nem todas as crianças inicialmente selecionadas compareceram a todas as avaliações planejadas. O número de indivíduos submetidos a cada teste foi, respectivamente, 3388, 2840 e 2420. Estes testes não foram planejados de forma a tornar comparáveis as escalas dos parâmetros dos itens e populacionais, ou seja, não havia itens comuns entre eles. Para permitir a equalização, dois outros testes secundários foram aplicados em grupos de 594 e 627 indivíduos, onde o primeiro continha itens aplicados nos testes de 1992 e 1993, enquanto o segundo continha itens dos testes de 1993 e 1994. Estes dois testes foram aplicados posteriormente aos três testes principais.

Para efeitos de estimação, serão considerados 5 testes, onde os testes principais serão denominados de testes 1, 2 e 3, enquanto os secundários serão os testes 4 e 5. O número de itens em cada teste são, respectivamente, 66, 48, 63, 66 e 70. O número de itens comuns é 38 entre os testes 1 e 4, 28 entre os testes 2 e 4, 28 entre os testes 2 e 5 e 42 entre os testes 3 e 5, perfazendo um total de $n = 177$ itens envolvidos no estudo. Devemos, entretanto, salientar que apenas os testes 1, 2 e 3 são de interesse, de forma que não apresentaremos os resultados referentes aos testes 4 e 5.

Como tanto os parâmetros populacionais como os dos itens são desconhecidos, será feita a estimação conjunta desses parâmetros. Tais estimações serão realizadas considerando algumas das estruturas de covariância exploradas neste trabalho. Serão apresentadas as estimativas de todos os parâmetros envolvidos no estudo, e para quantificar qual estrutura de covariância será a mais adequada, adotaremos o critério AIC (*Akaike's Information Criterion*), obtido por

$$AIC = -2 \log L(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\eta}}) + 2n_p,$$

onde a $L(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\eta}})$ representa o valor da verossimilhança, e n_p o número de parâmetros envolvidos no modelo (ver Jones, 1993, por exemplo). A estrutura de covariância que gerar o menor valor para AIC será a escolhida. O erro-padrão associado a cada parâmetro será estimado com base na inversa da matriz de informação de Fisher.

Na Tabela 9.11 temos as estimativas dos parâmetros populacionais associados aos três testes principais, quando adotados os modelos Uniforme, AR(1) e Hankel. O teste 1, relativo à primeira série, foi adotado como referência, de modo que a habilidade desta população no momento do teste 1 foi fixada como tendo média zero e desvio-padrão 1. As médias

Tabela 9.11 *Estimativas dos Parâmetros Populacionais (erro-padrão)*

	μ_1	μ_2	μ_3	ρ ou σ_{12}	σ_1^2	σ_2^2	σ_3^2
Uniforme	0,0000	0,8655 (0,0114)	1,5902 (0,0111)	0,8138 (0,0031)	1,0000	1,0000	1,0000
AR(1)	0,0000	0,8468 (0,0087)	1,6206 (0,0126)	0,9395 (0,0005)	1,0000	1,0000	1,0000
Hankel	0,0000	0,8770 (0,0189)	1,5927 (0,0196)	0,7219 (0,0117)	1,0000	0,8692 (0,0201)	1,0281 (0,0226)

populacionais se apresentaram bastante próximas para as três estruturas, indicando que na segunda série houve um aumento em torno de 0,85 pontos na habilidade média, quando comparada à primeira série. Para a terceira série, este aumento ficou em torno de 1,60. Entretanto, um dos principais resultados se refere à dependência entre as habilidades nas três séries. Quando considerada que a correlação entre as habilidades em qualquer par de testes é sempre a mesma [modelo Uniforme], a correlação estimada se apresentou bastante alta, ficando em 0,8138. Se considerarmos que a dependência diminui quando o tempo entre os testes aumenta [modelo AR(1)], a estimativa do parâmetro de interesse ficou em $\hat{\rho} = 0,9395$. Com esta estrutura temos que a correlação entre os testes 1 e 2 (e 2 e 3) é 0,9395, e entre os testes 1 e 3 é 0,8827. Se impormos as covariâncias todas iguais com as variâncias diferentes [modelo de Hankel], esta covariância será estimada em $\hat{\sigma}_{12} \simeq 0,7219$ e os desvios-padrão relativos às Séries 2 e 3 por 0,9323 ($= \sqrt{0,8692}$) e 1,0140 ($= \sqrt{1,0281}$), respectivamente, produzindo as correlações 0,7743, 0,7120 e 0,7637, correspondentes aos testes 1 e 2, 1 e 3, e 2 e 3.

Em todos os três casos vimos que as correlações estimadas foram altas, particularmente maior que 0,5, indicando um primeiro problema quando considerada a estrutura de Bandas. Entretanto, o maior problema foi sua não adaptação aos dados. Ela impõe que as correlações sejam nulas entre pares de testes não adjacentes. Ocorre que a correlação entre as habilidades nas Séries 1 e 3 são altas, digamos $\hat{\rho}_{13} \simeq 0,8138$ quando considerada a estrutura Uniforme, enquanto a estrutura de bandas impõe $\rho_{13} = 0$ e essa falta de ajuste aos dados foi refletida pela não convergência do processo iterativo. Ainda, foram analisados apenas os modelos que considerem algum tipo de dependência entre as habilidades, o que levou ao descarte do Modelo Diagonal.

Na Tabela 9.12 apresentamos as log-verossimilhanças, número de parâmetros e o valor para AIC quando adotados os modelos citados acima, que será usado para a decisão sobre qual dessas três estruturas de covariância melhor se adapta aos dados. A princípio, o fato de o modelo de Hankel ser uma generalização do caso Uniforme, deveríamos esperar que aquela estrutura fosse a eleita. Entretanto, aparentemente o fato de termos mais

Tabela 9.12 *Log-verossimilhança associada a cada estrutura de covariância*

Modelo	$-2 \times \text{Log-verossimilhança}$	Num. Parâmetros	AIC
Uniforme	513.920,00	536	514.992,00
AR(1)	517.940,00	536	519.012,00
Hankel	515.400,00	540	516.480,00

parâmetros a estimar (embora não tenha diferido muito da Uniforme) teve seu preço. Tivemos um menor valor de AIC quando adotado o modelo Uniforme, seguidos por Hankel e AR(1), respectivamente. Com base nisso, optamos por eleger o modelo Uniforme como o que melhor se ajustou aos dados, e que a correlação estimada entre as habilidades entre qualquer par de testes é $\hat{\rho} \simeq 0,8138$. Entretanto, apesar da estrutura Uniforme ter sido a escolhida, isso não significa que ela tenha se ajustado muito bem aos dados. Estudos no sentido de *Ajuste do Modelo* ainda devem ser feitos para chegar em resultados conclusivos.

Tabela 9.13 *Soma das variâncias associadas a cada parâmetro*

Modelo	a	b	c
Uniforme	0,920	9,898	0,619
AR(1)	0,970	10,833	0,691
Hankel	0,935	10,270	0,635

Como complemento, podemos usar a VT para inspecionar qual modelo produziu estimativas mais precisas dos parâmetros dos itens. Na Tabela 9.13 apresentamos as VTs associadas aos parâmetros de discriminação, dificuldade e acerto ao acaso, respectivamente. Novamente, a estrutura de correlação que produziu melhores resultados foi a Uniforme, levando-nos a reforçar a escolha do modelo Uniforme como o que melhor se adaptou aos dados. Nos resultados apresentados na Tabela 9.13 foram excluídos os resultados referentes a 5 dos 177 dos itens em estudo, por apresentarem estimação ruim nos 3 modelos em estudo. Entretanto, a conclusão foi inalterada pela retirada destes itens.

Comentários e Propostas para Pesquisas Futuras

Para finalizar, faremos alguns comentários sobre a proposta apresentada e teceremos outros sobre trabalhos que podem ser realizados como extensões deste, ou outros pontos não abordados aqui. Nesse conjunto de pontos são sugeridas linhas de trabalho bem gerais para serem desenvolvidas por futuros TRI-adeptos. Embora enumerados, não pretendemos atribuir prioridades ou níveis de importância a eles, mas apenas apresentá-los.

1. Modelos Multivariados

Em algumas situações, temos o mesmo indivíduo submetido a testes em diferentes conteúdos. Como Exemplos podemos indicar o Projeto da Avaliação da Jornada Única, discutido no Capítulo 9, e o Projeto PDE/Fundescola do INEP/MEC, nos quais os alunos são submetidos a testes de Português e Matemática em anos consecutivos. Neste caso, temos duas habilidades distintas envolvidas no estudo, com cada uma delas representadas por um único traço latente (unidimensional). Assim, a probabilidade de acerto por um determinado indivíduo j a um particular item i de Matemática pode ser modelada como função apenas da habilidade em Matemática do indivíduos j . Da mesma forma, a probabilidade de acerto pelo indivíduo j ao item i de Português depende apenas de sua habilidade em Português. A presença dessas duas habilidades distintas constitui o caso Multivariado destes modelos e tem muito a ser explorado. O fato de estarmos observando respostas do mesmo indivíduo, referente a duas habilidades, possibilita o estudo de possíveis correlações entre as mesmas de forma que possíveis estruturas de covariância podem ser propostas para modelar a dependência entre estas habilidades. Detalhes sobre Modelos Multivariados quando temos apenas uma condição de avaliação podem ser encontrados em Van der Linden & Hambleton (1997).

2. Distribuições assimétricas

Há situações em que não deve-se assumir uma distribuição simétrica para a distribuição da habilidade. Por exemplo, quando os indivíduos envolvidos no estudo foram pré-selecionados em uma primeira etapa, tal como ocorre em exames vestibulares. Nesta situação, devem ser propostas distribuições alternativas que não a Normal Multivariada e novas equações de estimação deverão ser desenvolvidas.

3. Estruturas de covariâncias

Embora tenhamos explorado cinco das principais estruturas de covariâncias, várias outras podem ser propostas com essa finalidade. Neste trabalho, usamos estruturas de covariâncias cujos determinantes e inversa tivessem uma forma analítica, de forma que os processos de estimação fossem construídos com base nas expressões exatas para as equações de estimação. Entretanto, grande parte dos programas estatísticos oferece a opção de usarmos derivadas numéricas, de forma que podemos estender este estudo para estruturas de covariâncias que não tenham expressões analíticas para as derivadas do determinante e da inversa. Um exemplo seria a matriz de covariância não estruturada, usualmente utilizada em análise multivariada. Estudos serão necessários para quantificar possíveis perdas da utilização de derivadas numéricas.

4. Modelos DRIFT para Genética

A Teoria da Resposta ao Item pode ser adaptada para aplicações em estudos genéticos. Consideremos a situação em que temos N indivíduos com uma determinada anomalia (como um particular tipo de câncer) envolvidos no estudo, dos quais são analisados n genes (itens) em T condições de avaliação com respostas *ativado/desativado*. A probabilidade do indivíduo j ter o gene i ativo no tempo t , fixada sua predisposição θ_{jt} (habilidade) é dada por

$$P(U_{jit} = 1 | \theta_{jt}, \boldsymbol{\zeta}_i) = c_i + (\gamma_i - c_i) \{1 + e^{-D a_i (\theta_{jt} - b_i)}\}^{-1},$$

onde $\boldsymbol{\zeta}_i = (a_i, b_i, c_i, \gamma_i)'$ é o vetor de parâmetros associado ao gene i ,

b_i é o parâmetro de inatividade (ou de posição) do gene i , medido na mesma escala da predisposição,

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

c_i é o parâmetro frequência positiva do gene i ,

γ_i é o parâmetro frequência negativa do gene i .

Esta função representa o Modelo Logístico de 4 parâmetros. No caso da aplicação DRIFT, o interesse se concentra no parâmetro de inatividade dos itens, de forma que os outros parâmetros podem ser considerados iguais, isto é, $a_i = a$, $c_i = c$ e $\gamma_i = \gamma$ para $i = 1, \dots, n$. Para cada gene, a tendência de alteração de sua inatividade pode ser modelada por uma função do tempo, tal como uma função linear ou quadrática, dando informação sobre o acréscimo ou decréscimo da influência de cada gene na anomalia estudada.

5. Parâmetro de Acerto Parcial

Se um particular indivíduo não responde a um ou mais itens e simplesmente ignorarmos estes itens na elaboração da verossimilhança, estaremos desprezando uma informação importante, a de que tal indivíduo não soube responder corretamente ao item. Se por outro lado estes itens forem tratados como respondidos incorretamente, estaremos desprezando outra informação importante, a de que este indivíduo poderia responder a esse item ao acaso, e com isso ele teria uma probabilidade positiva de respondê-lo corretamente. Uma proposta intermediária seria a *Imputação de Dados*, ou seja, a substituição de cada resposta incompleta por resposta correta, com probabilidade $\delta_i = 1/m_i$, ou resposta incorreta, com probabilidade $1 - \delta_i$, onde m_i é o número de alternativas do item i , e finalmente aplicar a teoria desenvolvida para dados completos; se os itens com respostas incompletas já tiverem sido calibrados (e portanto com os c_i conhecidos para o LM3) então pode-se usar c_i no lugar de δ_i . Lord (1974) argumenta que este procedimento é bastante criticável, mesmo produzindo estimadores consistentes ($n \rightarrow \infty$) das habilidades individuais. Como em situações práticas não é possível trabalhar com n muito grande, alguns indivíduos que tenham tido respostas imputadas podem ter suas habilidades mal estimadas devido ao conjunto de respostas geradas. Por outro lado, as estimativas dos parâmetros dos itens não devem ser fortemente alteradas porque geralmente o número de indivíduos, N , é relativamente alto. Para complementar, se dois indivíduos com habilidades θ_A e θ_B , com $\theta_A \neq \theta_B$, omitem a resposta ao item i , pela imputação de dados podemos ser levados a concluir que $P(U_{Ait} = 1 | \theta_A, \zeta_i) = \delta_i$ e $P(U_{Bit} = 1 | \theta_B, \zeta_i) = \delta_i$, e portanto que $\theta_A = \theta_B$, o que não é verdade. Lord também afirma que tal procedimento degrada os dados pela introdução de uma componente aleatória.

Como alternativa, Lord propõe uma modificação na função de verossimilhança que considere um item parcialmente correto, através da introdução de um novo parâmetro δ . A função de verossimilhança individual a tratar será exatamente a mesma em (6.3), com a alteração que as respostas serão codificadas como

$$U_{jit} = \begin{cases} 1 & : \text{resposta correta;} \\ 0 & : \text{resposta incorreta;} \\ \delta & : \text{resposta incompleta (parcialmente correta).} \end{cases} \quad (10.1)$$

No caso de não haverem observações incompletas, esta proposta coincide com o tratamento usual para dados completos. Lord enfatiza que seu método não produz estimativas de máxima verossimilhança na presença de observações incompletas, mas os valores se mostraram tão próximos que Wood, Wingersky & Lord (1976) tratam as estimativas obtidas em ambos os casos como estimativas de máxima verossimilhança. Lord utilizou dados simulados e verificou que as correlações entre as estimativas \hat{a}_i , \hat{b}_i e \hat{c}_i , obtidas pelos dois métodos foram 0.995, 0.996 e 0.990, respectivamente. Entretanto, o principal resultado de Lord foi a demonstração de que seu método produz uma menor variância

assintótica de $\hat{\theta}_j$, quando comparado ao estimador de máxima verossimilhança usual com os dados imputados.

6. Estimação Bayesiana dos Parâmetros Populacionais

A estimação por máxima verossimilhança irrestrita pode apresentar problemas na estimação de alguns parâmetros populacionais, devido a algumas restrições em seus espaços paramétricos. No modelo Diagonal devemos ter $\sigma_t > 0$, no modelo Uniforme, $\sigma^2 > 0$ e $\rho \in [-1, 1]$, no modelo de Bandas, $\sigma^2 > 0$ e $\rho \in [-\frac{1}{2}, \frac{1}{2}]$, no modelo AR(1), $\sigma^2 > 0$ e $\rho \in [-1, 1]$ e, finalmente, no de Hankel, $\sigma_t > 0$ e $|\sigma_{12}| < \min_{t,s} \{\sqrt{\sigma_t^2 \sigma_s^2}\}$. Para garantir que as estimativas permaneçam no intervalo desejado, distribuições *a priori* com suporte nestes intervalos devem ser adotadas para esses parâmetros. A importância dessas propostas é muito grande, pois é comum que os processos iterativos fiquem instáveis quando temos os parâmetros perto das bordas de seus intervalos, e esse é um caso bastante freqüente em estudos longitudinais.

7. Modelos Multidimensionais

Em algumas situações, um item não pode ser modelado como função apenas de um único traço latente. Alguns itens exigem, por exemplo, além do conhecimento em matemática, a habilidade em *compreensão de texto*; uma questão de química pode precisar, além do conhecimento em química, da habilidade em matemática para ser respondida corretamente. Itens dessa natureza precisam ser modelados considerando um vetor r -dimensional de habilidades $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_r)'$. Um dos modelos mais simples para modelar a probabilidade ao item i é dado por uma extensão do modelo logístico de 3 parâmetros (ver Reckase, 1997):

$$P(U_{jit} = 1 | \boldsymbol{\theta}_{jt}, \boldsymbol{\zeta}_i) = c_i + (1 - c_i) \{1 + e^{-D(\mathbf{a}'_i \boldsymbol{\theta}_{jt} - d_i)}\}^{-1},$$

onde \mathbf{a}_i é o vetor de parâmetros de discriminação do item i e, geometricamente, a_{il} representa o coeficiente angular da superfície de resposta na direção do eixo l da habilidade. A quantidade d_i passa a ser interpretada como o novo parâmetro de dificuldade do item.

8. Respostas Politômicas

Em algumas aplicações, as respostas dos indivíduos aos itens não podem e/ou não devem ser interpretadas como do tipo certo/errado (ver Andrade, Tavares & Valle, 2000, por exemplo). Nestes casos, os itens devem ser corrigidos como tendo mais de duas categorias de resposta e são denominados itens *politômicos*. No presente trabalho, apresentamos apenas as expressões para itens dicotômicos e, portanto, muito precisa ser feito no caso de termos itens com respostas politômicas, contavelmente finitas ou contínuas.

9. Estimação de Densidades

A estimação da densidade é um ponto de muito interesse nessa teoria. Na maioria das situações pode-se assumir que a distribuição da habilidade é Normal, e portanto simétrica, mas em outras a suposição de simetria é bastante irreal, conforme mencionado na *Seção 2*. Como a estimação dos vetores de habilidades individuais é feita em uma segunda etapa e é obtida com base na distribuição condicional da habilidade, dado o conjunto de respostas do indivíduo, os parâmetros dos itens e os populacionais, fugas da verdadeira distribuição da habilidade podem levar a erros grosseiros na estimação das habilidades individuais.

10. Estudos de Ajuste do Modelo, Convergência do EMV e Possíveis Correções

A validade do modelo precisa ser checada. Além disso, o número de indivíduos e o número de condições de avaliação, bem como o número de itens comuns, devem influenciar bastante na taxa de convergência para distribuição assintótica do EMV. Esses fatores precisam ser melhor estudados e possíveis correções decorrentes desses resultados podem ser propostas.

11. O Programa Computacional

Até o presente momento o programa computacional desenvolvido abrange somente parte da metodologia proposta nesse trabalho. Mais especificamente, o programa não abrange o caso de duas ou mais populações e nem o ajuste de curvas de crescimento para modelar o vetor de médias ao longo do tempo. O programa também não possui uma forma amigável de interação com o usuário. Devido a importância da metodologia proposta, faz-se necessária a implementação desses modelos no programa, bem como torná-lo disponível em uma versão beta para que ele possa ser testado por vários usuários. O objetivo final seria o desenvolvimento de uma versão comercial.

Desenvolvimentos Complementares: Matriz Hessiana

As equações de estimação para $\zeta_i = (a_i, b_i, c_i)'$, $i = 1, \dots, n$, e η não apresentam soluções explícitas para os estimadores de máxima verossimilhança dos parâmetros de interesse. Por conta disso, será necessário utilizarmos algum processo iterativo para encontrar as estimativas desejadas. O procedimento padrão é expandir as equações de estimação em série de Taylor até segunda ordem, usar uma estimativa inicial dos parâmetros e encontrar uma nova estimativa. Este processo é repetido até que algum critério de parada seja alcançado, e é conhecido como *Algoritmo Newton-Raphson* (Issac & Keller, 1966).

Uma modificação do algoritmo Newton-Raphson consiste em substituir as derivadas segundas da log-verossimilhança pelos seus valores esperados. Este processo é conhecido como *Método "Scoring" de Fisher* (Rao, 1973). Neste trabalho será usado o processo iterativo Newton-Raphson, pois este tem apresentado resultados satisfatórios, além de sua simplicidade.

Como a expansão será feita até segunda ordem, serão necessárias as derivadas segundas da log-verossimilhança com relação aos parâmetros dos itens, ζ_i , $i = 1, \dots, n$, parâmetros populacionais, η , e com relação a ambos. Como comentado nos Capítulos 2 e 3, as equações de estimação lá obtidas formam os dois blocos básicos para a estimação conjunta, mas também servem para a situação em que desejamos estimar os parâmetros dos itens com os parâmetros populacionais conhecidos (Capítulo 2), bem como estimar os parâmetros populacionais com os parâmetros dos itens conhecidos (Capítulo 4). Por conta disso, vamos apresentar em separado as derivadas segundas com relação aos parâmetros dos itens (Seção A.1) e parâmetros populacionais (Seção A.2). O caso conjunto é deixado para a Seção A.3. A Seção A.4 é destinada às habilidades, a Seção A.5 considera várias populações, a Seção A.6 trata de Curvas de Crescimento, com expressões para o caso geral, e a Seção A.7 destina-se à curva de Heitjan.

A.1 Parâmetros dos itens

Precisamos obter a derivada segunda de $\log L(\zeta, \eta)$ com relação a ζ_i e a ζ_l . De (2.9) temos que

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

onde a segunda parcela de (A.1) é obtida de (2.18). Com relação a primeira, seja

$$V_{jit} = \frac{(-1)^{U_{jit}+1}}{P(U_{jit} | \theta_t, \boldsymbol{\zeta}_i)} \quad \text{com} \quad P(U_{jit} | \theta_t, \boldsymbol{\zeta}_i) = P_{it}^{U_{jit}} Q_{it}^{1-U_{jit}},$$

e notemos que

$$\begin{aligned}
\frac{\partial V_{jit}}{\partial \boldsymbol{\zeta}_i} &= (-1)^{U_{jit}+1} \frac{-1}{P^2(U_{jit} | \theta_t, \boldsymbol{\zeta}_i)} \frac{\partial P(U_{jit} | \theta_t, \boldsymbol{\zeta}_i)}{\partial \boldsymbol{\zeta}_i} \\
&= -V_{jit} \frac{\partial P(U_{jit} | \theta_t, \boldsymbol{\zeta}_i) / \partial \boldsymbol{\zeta}_i}{P(U_{jit} | \theta_t, \boldsymbol{\zeta}_i)} \\
&= \frac{-V_{jit}}{P(U_{jit} | \theta_t, \boldsymbol{\zeta}_i)} (-1)^{U_{jit}+1} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \\
&= -V_{jit}^2 \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right),
\end{aligned} \tag{A.2}$$

onde a penúltima igualdade segue por (2.13). Vamos, agora, considerar em separado os casos $i = l$ e $i \neq l$.

Caso $i = l$: De (2.15) e (A.2) temos que

$$\begin{aligned}
\frac{\partial^2 P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}'_i} &= \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left\{ \int_{\mathbb{R}^T} \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \right] P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta} \right\}' \\
&= \int_{\mathbb{R}^T} \left\{ \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right)' \right] P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) + \right. \\
&\quad \left. + \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \right] \left[\frac{\partial P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} \right]' \right\} g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}.
\end{aligned} \tag{A.3}$$

Notemos agora que

$$\frac{\partial}{\partial \zeta_i} \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial}{\partial \zeta_i} P_{it} \right)' \right] = \sum_{t \in \tau_i} V_{jit} \left(\frac{\partial^2 P_{it}}{\partial \zeta_i \partial \zeta_i'} \right) - \sum_{t \in \tau_i} V_{jit}^2 \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) \left(\frac{\partial P_{it}}{\partial \zeta_i} \right)'$$

e, por (2.11) e (2.18),

$$\frac{\partial P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\eta})}{\partial \zeta_i} = \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial P_{it}}{\partial \zeta_i} \right)' \right] P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}).$$

Segue com isso que (A.3) pode ser escrita como

$$\begin{aligned} \frac{\partial^2 P(\mathbf{U}_{j..} | \boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i \partial \zeta_i'} &= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} V_{jit} \left(\frac{\partial^2 P_{it}}{\partial \zeta_i \partial \zeta_i'} \right) - \sum_{t \in \tau_i} V_{jit}^2 \left(\frac{\partial P_{it}}{\partial \zeta_i} \right) \left(\frac{\partial P_{it}}{\partial \zeta_i} \right)' + \right. \\ &\quad \left. + \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial P_{it}}{\partial \zeta_i} \right)' \right] \left[\sum_{t \in \tau_i} V_{jit} \left(\frac{\partial P_{it}}{\partial \zeta_i} \right)' \right]' \right\} P(\mathbf{U}_{j..} | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}. \quad (\text{A.4}) \end{aligned}$$

Vale notar que de (2.14) obtemos

$$V_{jit} = \frac{U_{jit} - P_{it}}{P_{it} Q_{it}} = (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^* Q_{it}^*}.$$

Podemos agora obter as equações para as componentes de ζ_i . Antes, precisaremos das derivadas segundas de P_{it} com relação a a_i , b_i e c_i , que são as seguintes:

$$\begin{aligned} \frac{\partial P_{it}^* Q_{it}^*}{\partial a_i} &= (1 - 2P_{it}^*) \frac{\partial P_{it}}{\partial a_i}, \quad \alpha_i \in \{a_i, b_i, c_i\} \\ \frac{\partial^2 P_{it}}{\partial a_i^2} &= D^2(1 - c_i)(\theta_t - b_i)^2 P_{it}^* Q_{it}^* (1 - 2P_{it}^*) \end{aligned} \quad (\text{A.5})$$

$$\frac{\partial^2 P_{it}}{\partial a_i \partial b_i} = -D(1 - c_i) P_{it}^* Q_{it}^* \{1 + Da_i(\theta_t - b_i)(1 - 2P_{it}^*)\} \quad (\text{A.6})$$

$$\frac{\partial^2 P_{it}}{\partial a_i \partial c_i} = -D(\theta_t - b_i) P_{it}^* Q_{it}^* \quad (\text{A.7})$$

$$\frac{\partial^2 P_{it}}{\partial b_i^2} = D^2 a_i^2 (1 - c_i) P_{it}^* Q_{it}^* (1 - 2P_{it}^*) \quad (\text{A.8})$$

$$\frac{\partial^2 P_{it}}{\partial b_i \partial c_i} = Da_i P_{it}^* Q_{it}^* \quad (\text{A.9})$$

$$\frac{\partial^2 P_{it}}{\partial c_i^2} = \frac{\partial Q_{it}^*}{\partial c_i} = 0 \quad (\text{A.10})$$

Com estas expressões obtemos $\partial^2 P_{it}/\partial\zeta_i\partial\zeta'_i$. Sejam

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

Retornando a (A.4), teremos que

$$\begin{aligned} \mathbf{H}_{ii(j)} &\equiv \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})/\partial\zeta_i\partial\zeta'_i}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \\ &= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{H}_{iit} - \sum_{t \in \tau_i} (U_{jit} - P_{it})^2 W_{it}^2 \mathbf{h}_{it} \mathbf{h}'_{it} + \right. \\ &\quad \left. + \left[\sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right] \left[\sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}'_{it} \right] \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Caso $i \neq l$: De forma similar ao caso anterior, temos

$$\begin{aligned} \mathbf{H}_{il(j)} &\equiv \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})/\partial\zeta_i\partial\zeta'_l}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \\ &= \int_{\mathbb{R}^T} \left[\sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right] \left[\sum_{t \in \tau_l} (U_{jlt} - P_{lt}) W_{lt} \mathbf{h}'_{lt} \right] g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Com as notações acima, da expressão (2.18) podemos escrever

$$\mathbf{h}_{i(j)} \equiv \frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})/\partial\zeta_i}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Retornando a (A.1), obtemos

$$\begin{aligned}
\mathbf{H}(\zeta_i, \zeta_l) &= \frac{\partial^2 \log L(\zeta, \eta)}{\partial \zeta_i \partial \zeta_l'} \\
&= \sum_{j=1}^s r_j \{ \mathbf{H}_{il(j)} - \mathbf{h}_{i(j)} \mathbf{h}_{l(j)}' \}.
\end{aligned} \tag{A.11}$$

Considerando

$$\begin{aligned}
\mathbf{f}(\zeta_i) &= \frac{\partial \log L(\zeta, \eta)}{\partial \zeta_i} \\
&= \sum_{j=1}^s r_j \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) W_{it} \mathbf{h}_{it} \right\} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}$$

bem como

$$\mathbf{f}_I(\zeta) = \begin{pmatrix} \mathbf{f}(\zeta_1) \\ \vdots \\ \mathbf{f}(\zeta_n) \end{pmatrix} \quad \text{e} \quad \mathbf{H}_I(\zeta) = \begin{pmatrix} \mathbf{H}(\zeta_1, \zeta_1) & \cdots & \mathbf{H}(\zeta_1, \zeta_n) \\ \vdots & \ddots & \vdots \\ \mathbf{H}(\zeta_n, \zeta_1) & \cdots & \mathbf{H}(\zeta_n, \zeta_n) \end{pmatrix}.$$

temos que se $\widehat{\zeta}^{(k)}$ é uma estimativa de ζ na iteração k , então na iteração $k + 1$ teremos que

$$\widehat{\zeta}^{(k+1)} = \widehat{\zeta}^{(k)} - [\mathbf{H}_I(\widehat{\zeta}^{(k)})]^{-1} \mathbf{f}_I(\widehat{\zeta}^{(k)}).$$

A.2 Parâmetros populacionais

Por (3.2) a derivada segunda da log-verossimilhança é

$$\begin{aligned}
\frac{\partial^2 \log L(\zeta, \eta)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} &= \\
&= \sum_{j=1}^s r_j \left\{ \frac{\partial^2 P(\mathbf{U}_{j..} | \zeta, \eta) / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}{P(\mathbf{U}_{j..} | \zeta, \eta)} - \left(\frac{\partial P(\mathbf{U}_{j..} | \zeta, \eta) / \partial \boldsymbol{\eta}}{P(\mathbf{U}_{j..} | \zeta, \eta)} \right) \left(\frac{\partial P(\mathbf{U}_{j..} | \zeta, \eta) / \partial \boldsymbol{\eta}}{P(\mathbf{U}_{j..} | \zeta, \eta)} \right)' \right\},
\end{aligned} \tag{A.12}$$

onde a derivada no segundo termo é dada por (3.3). Para a primeira parcela, notemos que

$$\begin{aligned}
\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} &= \\
&= \frac{\partial}{\partial \boldsymbol{\eta}} \left\{ \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right) d\boldsymbol{\theta} \right\} \\
&= \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left[\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} + \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right) \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)' \right] d\boldsymbol{\theta}.
\end{aligned}$$

Portanto, precisaremos das derivadas segundas de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta})$. Para o caso Normal, temos

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_t \partial \mu_s} = -v_{t,s}, \quad \text{para } t, s = 1, \dots, T.$$

Em notação matricial, temos

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}'} = -\mathbf{V}. \quad (\text{A.13})$$

De forma geral,

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \sigma_{ij}} = \left[\frac{\partial \mathbf{V}}{\partial \sigma_{ij}} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}) \quad (\text{A.14})$$

e

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \sigma_{ij} \partial \sigma_{i'j'}} = -\frac{1}{2} \left[\frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \sigma_{ij} \partial \sigma_{i'j'}} \right] - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \left[\frac{\partial^2 \mathbf{V}}{\partial \sigma_{ij} \partial \sigma_{i'j'}} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (\text{A.15})$$

Sejam

$$\mathbf{f}_P(\boldsymbol{\eta}) = \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}, \quad \mathbf{H}_P(\boldsymbol{\eta}) = \frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}.$$

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

$$\hat{\boldsymbol{\eta}}^{(k+1)} = \hat{\boldsymbol{\eta}}^{(k)} - [\mathbf{H}_P(\hat{\boldsymbol{\eta}}^{(k)})]^{-1} \mathbf{f}_P(\hat{\boldsymbol{\eta}}^{(k)}). \quad (\text{A.16})$$

As expressões exatas necessárias ao processo de estimação dependem da estrutura de covariância adotada. A seguir, aplicaremos o procedimento Newton-Raphson para estimação de cada uma das estruturas de covariâncias consideradas. Basicamente, o que faremos é obter as expressões das derivadas segundas da log-verossimilhança com relação aos parâmetros populacionais. Com estas expressões substituídas em (A.12) e aplicadas em (A.16) podemos chegar às estimativas procuradas.

Matriz Diagonal

As expressões para as derivadas segundas são

$$\frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial(\sigma_t^2)^2} = -\frac{1}{\sigma_t^4},$$

$$\mathbf{V}_t^{(2)} = \left(\frac{\partial^2 v_{ij}}{\partial(\sigma_t^2)^2} \right),$$

onde

$$\frac{\partial v_{ij}}{\partial \sigma_t^2} = \begin{cases} 2/\sigma_t^6 & \text{se } i = j = t ; \\ 0 & \text{caso contrário.} \end{cases}$$

As derivadas com relação a σ_t^2 e σ_s^2 são nulas se $t \neq s$. Ainda,

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \mu_s \partial \sigma_t^2} = \begin{cases} -(\theta_t - \mu_t)/\sigma_t^4 & \text{se } t = s \\ 0 & \text{c.c.} \end{cases} \quad (\text{A.17})$$

Matriz Uniforme

As expressões para as derivadas segundas são:

$$\frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \rho^2} = -(T-1) \left[\frac{1}{\alpha^2} + \frac{T-1}{\beta^2} \right]$$

$$\mathbf{V}_\rho^{(2)} \equiv \frac{\partial^2 \mathbf{V}}{\partial \rho^2} = \frac{2}{\alpha^3 \sigma^2} \mathbf{I}_T + \frac{2}{(\alpha\beta)^3 \sigma^2} ((\delta - \rho(T-1))\alpha\beta - \rho\delta^2) \mathbf{J}_T.$$

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \rho} = \left[\frac{\partial \mathbf{V}}{\partial \rho} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}) = \mathbf{V}_\rho^{(1)} (\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (\text{A.18})$$

Para as matrizes Uniforme, Bandas e AR(1), as derivadas com relação a σ^2 são

$$\frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial(\sigma^2)^2} = -T\sigma^{-4}, \quad \frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \sigma^2 \partial \rho} = 0, \quad (\text{A.19})$$

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

Temos ainda que

$$\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \sigma^2} = \left[\frac{\partial \mathbf{V}}{\partial \sigma^2} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}) = -\sigma^{-2} \mathbf{V} (\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (\text{A.20})$$

Matriz de Bandas

Neste caso, temos que

$$\frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \rho^2} = - \sum_{t=1}^T \left\{ \rho + \frac{1}{2} \cos^{-1} \left(\frac{t\pi}{T+1} \right) \right\}^{-2},$$

$$\mathbf{V}^{(2)} = (\sigma^{-2} v_{ij}^{(2)}),$$

com

$$v_{ij}^{(2)} = \frac{\partial^2 v_{ij}}{\partial \rho} = \frac{\gamma'_{4ij} \rho^2 \gamma_3^2 - 2\rho \gamma_3 \gamma'_{4ij} (\gamma_3 + \rho \gamma_3')}{(\rho \gamma_3)^4},$$

onde

$$\gamma'_{4ij} = \rho \gamma_3 (\gamma''_{1ij} \gamma_{2ij} + 2\gamma'_{1ij} \gamma'_{2ij} + \gamma_{1ij} \gamma''_{2ij}) - \gamma_{1ij} \gamma_{2ij} (2\gamma_3' + \rho \gamma_3'').$$

As derivadas segundas de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta})$ com relação a $\boldsymbol{\mu}$ e ρ são dadas por (A.18), e com relação a σ^2 por (A.19) a (A.20).

Matriz AR(1)

Neste caso teremos que

$$\frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \rho^2} = -2(T-1) \left[\frac{1 + \rho^2}{(1 - \rho^2)^2} \right].$$

$$\mathbf{V}^{(2)} = (\sigma^{-2} v_{ij}^{(2)}),$$

com

$$v_{ij}^{(2)} = \begin{cases} \frac{2(\alpha+4\rho^2)}{\alpha^3} & \text{se } i = j \text{ e } i \in \{1, T\} \\ \frac{4(\alpha+4\rho^2)}{\alpha^3} & \text{se } j \in \{2, 3, \dots, T-1\} \\ \frac{-2\rho(\alpha^2+\beta)}{\alpha^4} & \text{se } i = j \text{ e } i \in \{i-1, i+1\} \\ 0 & \text{c.c.} \end{cases}$$

As derivadas segundas de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta})$ com relação a $\boldsymbol{\mu}$ e ρ são dadas por (A.18), e com relação a σ^2 por (A.19) a (A.20).

Matriz Hankel

Neste caso temos que $\boldsymbol{\eta} = (\mu_1, \mu_2, \dots, \mu_T, \sigma_{12}, \sigma_{11}, \sigma_{22}, \dots, \sigma_{TT})'$. As derivadas segundas do $\log |\boldsymbol{\Sigma}|$ são

$$\begin{aligned}\frac{\partial^2}{\partial \sigma_{12}^2} \log |\boldsymbol{\Sigma}| &= \frac{2\alpha\alpha_2 - \alpha_1^2}{\alpha^2} - S_2; \\ \frac{\partial^2}{\partial \sigma_{ii}^2} \log |\boldsymbol{\Sigma}| &= -\frac{\gamma(\gamma + 2d_i)}{d_i^4} - \frac{1}{d_i^2}, \quad i = 1, \dots, T; \\ \frac{\partial^2}{\partial \sigma_{12} \partial \sigma_{ii}} \log |\boldsymbol{\Sigma}| &= \frac{\delta d_i - 2\sigma_{12}\alpha}{\alpha^2 d_i^3} + \frac{1}{d_i^2}, \quad i = 1, \dots, T; \\ \frac{\partial^2}{\partial \sigma_{ii} \partial \sigma_{jj}} \log |\boldsymbol{\Sigma}| &= \frac{-\gamma^2}{(d_i d_j)^2}, \quad \text{para } i \neq j.\end{aligned}$$

As derivadas segundas de \mathbf{V} com relação a σ_{12} e σ_{ii} são, respectivamente,

$$\begin{aligned}\mathbf{V}_{12}^{(2)} &= \frac{\partial^2 \mathbf{V}}{\partial \sigma_{12}^2} = \mathbf{F}_{12}^{(2)} + \gamma_{12}^{(2)} \mathbf{E} + 2\gamma_{12}^{(1)} \mathbf{E}_{12}^{(1)} + \gamma \mathbf{E}_{12}^{(2)}; \\ \mathbf{V}_{ii}^{(2)} &= \frac{\partial^2 \mathbf{V}}{\partial \sigma_{ii}^2} = \mathbf{F}_{ii}^{(2)} + \gamma_{ii}^{(2)} \mathbf{E} + 2\gamma_{ii}^{(1)} \mathbf{E}_{ii}^{(1)} + \gamma \mathbf{E}_{ii}^{(2)}, \quad i = 1, \dots, T. \\ \mathbf{V}_{12,ii}^{(2)} &= \frac{\partial^2 \mathbf{V}}{\partial \sigma_{12} \partial \sigma_{ii}} = \mathbf{F}_{12,ii}^{(2)} + \gamma_{12,ii}^{(2)} \mathbf{E} + \gamma_{12}^{(1)} \mathbf{E}_{ii}^{(1)} + \gamma_{ii}^{(1)} \mathbf{E}_{12}^{(1)} + \gamma \mathbf{E}_{12,ii}^{(2)}, \quad i = 1, \dots, T. \\ \mathbf{V}_{ii,jj}^{(2)} &= \frac{\partial^2 \mathbf{V}}{\partial \sigma_{ii} \partial \sigma_{jj}} = \gamma_{ii,jj}^{(2)} \mathbf{E} + \gamma_{ii}^{(1)} \mathbf{E}_{jj}^{(1)} + \gamma_{jj}^{(1)} \mathbf{E}_{ii}^{(1)} + \gamma \mathbf{E}_{ii,jj}^{(2)}, \quad i \neq j.\end{aligned}$$

Resta agora obtermos as expressões para as derivadas de γ , \mathbf{E} e \mathbf{F} . As derivadas segundas de γ são

$$\begin{aligned}\gamma_{12}^{(2)} &= \frac{2(\sigma_{12}\alpha_2\alpha - \alpha_1\delta)}{\alpha^4}; \\ \gamma_{ii}^{(2)} &= \frac{2\gamma^2(\gamma + d_i)}{d_i^4}, \quad i = 1, \dots, T \\ \gamma_{12,ii}^{(2)} &= \frac{-2(\gamma\delta d_i + \sigma_{12}^2)}{\alpha^2 d_i^3}, \quad i = 1, \dots, T \\ \gamma_{ii,jj}^{(2)} &= \frac{2\gamma^3}{(d_i d_j)^2}, \quad i \neq j.\end{aligned}$$

As derivadas segundas de \mathbf{F} são

$$f_{ts,12}^{(2)} = \begin{cases} 2/d_i^3 & \text{se } t = s = i \\ 0 & \text{c.c.} \end{cases}$$

$$f_{ts,ii}^{(2)} = \begin{cases} 2/d_i^3 & \text{se } t = s = i \\ 0 & \text{c.c.} \end{cases}$$

$$f_{ts,12,ii}^{(2)} = \begin{cases} -2/d_i^3 & \text{se } t = s = i \\ 0 & \text{c.c.} \end{cases}$$

As derivadas segundas de \mathbf{E} são

$$e_{ij,12}^{(2)} = \frac{2(d_i + d_j)^2 - 2d_i d_j}{(d_i d_j)^3},$$

$$e_{ij,tt}^{(2)} = \begin{cases} 2/(d_i^3 d_j) & \text{se } t = i, i \neq j \\ 6/d_i^4 & \text{se } t = i = j \\ 0 & \text{c.c.} \end{cases}$$

$$e_{ij,12,tt}^{(2)} = \begin{cases} -(d_i + 2d_j)/(d_i^3 d_j^2) & \text{se } t = i, i \neq j \\ -6/d_i^4 & \text{se } t = i = j \\ 0 & \text{c.c.} \end{cases}$$

$$e_{ij,tt,ss}^{(2)} = \begin{cases} 6/d_i^4 & \text{se } t = i, s = j, i \neq j \\ 1/(d_i d_j) & \text{se } t = s = i = j \\ 0 & \text{c.c.} \end{cases}$$

Temos ainda que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \sigma_{12}} = \mathbf{V}_{12}^{(1)}(\boldsymbol{\theta} - \boldsymbol{\mu}),$$

onde $\mathbf{V}_{12}^{(1)}$ é a matriz $\partial \mathbf{V} / \partial \sigma_{12}$. Denotando $\boldsymbol{\eta}_2 = (\sigma_{11}, \sigma_{22}, \sigma_{33}, \dots, \sigma_{TT})'$, teremos que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\eta}'_2} = \mathbf{diag}(v_{11}(\theta_1 - \mu_1), v_{22}(\theta_2 - \mu_2), \dots, v_{TT}(\theta_T - \mu_T)).$$

A.3 Parâmetros dos Itens e Parâmetros Populacionais

Em notação matricial, temos que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu}_{(1)} \partial \boldsymbol{\mu}'_{(1)}} = -\mathbf{V}_{(1,1)}, \quad (\text{A.21})$$

onde $\mathbf{V}_{(1,1)}$ é a matriz \mathbf{V} sem a primeira linha e a primeira coluna. Teremos ainda que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu}_{(1)} \partial \sigma_{ij}} = \left[\frac{\partial \mathbf{V}_{(1)}}{\partial \sigma_{ij}} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (\text{A.22})$$

Da equação (3.2) segue que para $i = 1, \dots, n$,

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'} &= \\ &= \sum_{j=1}^s r_j \left\{ \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) / \partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} - \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) / \partial \boldsymbol{\zeta}_i}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \right) \left(\frac{\partial P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} \right)' \right\}. \end{aligned} \quad (\text{A.23})$$

Os elementos da segunda parcela de (A.23) são obtidos por (2.18) e (3.3), respectivamente. Com relação à primeira parcela, de (3.4) temos que

$$\begin{aligned} \frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'} &= \\ &= \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left\{ \int_{\mathbb{R}^T} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right)' d\boldsymbol{\theta} \right\} \\ &= \int_{\mathbb{R}^T} \left(\frac{\partial}{\partial \boldsymbol{\zeta}_i} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) \right) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right)' d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^* Q_{it}^*} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \right\} P(\mathbf{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}) \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right)' d\boldsymbol{\theta}. \end{aligned}$$

Segue que a primeira parcela de (A.23) pode ser escrita como

$$\frac{\partial^2 P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta}) / \partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'}{P(\mathbf{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\eta})} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{jit} - P_{it}) \frac{W_{it}}{P_{it}^* Q_{it}^*} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \right\} \left(\frac{\partial}{\partial \boldsymbol{\eta}} \log g(\boldsymbol{\theta}|\boldsymbol{\eta}) \right)' g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Sejam

$$\mathbf{H}_{iP}(\boldsymbol{\zeta}_i, \boldsymbol{\eta}) = \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'}, \quad \mathbf{H}_{*P}(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \begin{pmatrix} \mathbf{H}_{1P}(\boldsymbol{\zeta}_1, \boldsymbol{\eta}) \\ \vdots \\ \mathbf{H}_{nP}(\boldsymbol{\zeta}_n, \boldsymbol{\eta}) \end{pmatrix}.$$

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

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

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

$$\widehat{\boldsymbol{\psi}}^{(k+1)} = \widehat{\boldsymbol{\psi}}^{(k)} - [\mathbf{H}_{IP}(\widehat{\boldsymbol{\psi}}^{(k)})]^{-1} \mathbf{f}_{IP}(\widehat{\boldsymbol{\psi}}^{(k)}).$$

A.3.1 Expressões para o Algoritmo EM

Nesta seção obtemos as expressões para aplicação do algoritmo EM. Consideramos fixas as quantidades $\mathbf{f} = (f(\boldsymbol{\theta}))$, $\mathbf{r} = (r_{it}(\boldsymbol{\theta}))$ e os parâmetros populacionais. Com a utilização da independência entre os itens, precisaremos apenas das derivadas segundas da log-verossimilhança com relação a $\boldsymbol{\zeta}_i$, $i = 1, \dots, n$. De (4.6) temos que

$$\begin{aligned} \mathbf{H}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_i) &= \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} \frac{\partial}{\partial \boldsymbol{\zeta}_i} \left[(r_{it} - P_{it} f_{it}) \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \frac{W_{it}}{P_{it}^* Q_{it}^*} \right] \right\} d\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^T} \sum_{t \in \tau_i} W_{it} \left\{ \left[r_{it} - f_{it} \left(\frac{\partial P_{it}}{\partial \boldsymbol{\zeta}_i} \right) \right] \mathbf{h}'_{it} + [r_{it} - P_{it} f_{it}] [\mathbf{H}_{iit} - W_{it} (1 - 2P_{it}) \mathbf{h}_{it} \mathbf{h}'_{it}] \right\} d\boldsymbol{\theta}. \end{aligned}$$

A.4 Habilidades

A Equação (5.15) não possui solução explícita, logo precisaremos aplicar algum processo iterativo. Para aplicação do procedimento Newton-Raphson será necessária a derivada segunda de $g_j^*(\boldsymbol{\theta}_j)$. Temos que

$$\frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} = \frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} + \frac{\partial^2 \log g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'}. \quad (\text{A.24})$$

De (5.16), segue que

$$\begin{aligned} \frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \boldsymbol{\theta}_{jt}^2} &= \frac{\partial}{\partial \boldsymbol{\theta}_{jt}} \left(\sum_{i \in \mathbf{I}_t} \alpha_i V_{jit} P_{jit}^* Q_{jit}^* \right) \\ &= \sum_{i \in \mathbf{I}_t} \alpha_i \left[\left(\frac{\partial V_{jit}}{\partial \boldsymbol{\theta}_{jt}} \right) P_{jit}^* Q_{jit}^* + \left(\frac{\partial P_{jit}^* Q_{jit}^*}{\partial \boldsymbol{\theta}_{jt}} \right) V_{jit} \right]. \end{aligned}$$

Mas,

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

Com isso,

$$\begin{aligned}\frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \theta_{jt}^2} &= \sum_{i \in \mathbf{I}_t} \alpha_i \left[-V_{jit}^2 \left(\frac{\partial P_{jit}}{\partial \theta_{jt}} \right) P_{jit}^* Q_{jit}^* + (1 - 2P_{jit}^*) \left(\frac{\partial P_{jit}}{\partial \theta_{jt}} \right) V_{jit} \right] \\ &= \sum_{i \in \mathbf{I}_t} \alpha_i \left[-V_{jit}^2 P_{jit}^* Q_{jit}^* + (1 - 2P_{jit}^*) V_{jit} \right] \left(\frac{\partial P_{jit}}{\partial \theta_{jt}} \right) \\ &= \sum_{i \in \mathbf{I}_t} \alpha_i^2 V_{jit} \left[-V_{jit} P_{jit}^* Q_{jit}^* + (1 - 2P_{jit}^*) \right] P_{jit}^* Q_{jit}^*.\end{aligned}\quad (\text{A.25})$$

Temos ainda que

$$\frac{\partial^2 \log P(\mathbf{U}_{j..} | \boldsymbol{\theta}_j, \boldsymbol{\zeta})}{\partial \theta_{jt} \partial \theta_{js}} = 0, \quad \forall t \neq s.$$

Também, de (5.17),

$$\frac{\partial^2 \log g(\boldsymbol{\theta}_j | \boldsymbol{\eta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} = -\frac{\partial}{\partial \boldsymbol{\theta}_j} \left[\boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_j - \boldsymbol{\mu}) \right]' = -\boldsymbol{\Sigma}^{-1}.\quad (\text{A.26})$$

Com as expressões (A.25) e (A.26) chegamos a (A.24). Sejam

$$\mathbf{f}_\theta(\boldsymbol{\theta}_j) = \frac{\partial \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}, \quad \mathbf{H}_\theta(\boldsymbol{\theta}_j) = \frac{\partial^2 \log g_j^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'}\quad (\text{A.27})$$

Se $\widehat{\boldsymbol{\theta}}_j^{(k)}$ é uma estimativa de $\boldsymbol{\theta}_j$ na iteração k , então na iteração $k + 1$ teremos que

$$\widehat{\boldsymbol{\theta}}_j^{(k+1)} = \widehat{\boldsymbol{\theta}}_j^{(k)} - [\mathbf{H}_\theta(\widehat{\boldsymbol{\theta}}_j^{(k)})]^{-1} \mathbf{f}_\theta(\widehat{\boldsymbol{\theta}}_j^{(k)}).$$

A.5 Parâmetros dos Itens e Parâmetros Populacionais: Várias Populações

As expressões para o caso de várias populações são bastante similares às apresentadas na Seção A.3. Por isso, apenas as expressões principais serão apresentadas.

A.5 Parâmetros dos Itens e Parâmetros Populacionais: Várias Populações 97

Em notação matricial, temos que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\mu}_{k(1)} \partial \boldsymbol{\mu}'_{k(1)}} = -\mathbf{V}_{k(1,1)}, \quad (\text{A.28})$$

onde $\mathbf{V}_{k(1,1)}$ é a matriz \mathbf{V}_k sem a primeira linha e a primeira coluna. Teremos ainda que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta})}{\partial \boldsymbol{\mu}_{k(1)} \partial \sigma_{kij}} = \left[\frac{\partial \mathbf{V}_{k(1)}}{\partial \sigma_{kij}} \right] (\boldsymbol{\theta} - \boldsymbol{\mu}_k). \quad (\text{A.29})$$

Da equação (3.2) segue que para $i = 1, \dots, n$,

$$\begin{aligned} & \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'_k} = \\ & = \sum_{k=1}^K \sum_{j=1}^{N_k} \left\{ \frac{\partial^2 P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k) / \partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'_k}{P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} - \left(\frac{\partial P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k) / \partial \boldsymbol{\zeta}_i}{P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} \right) \left(\frac{\partial P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k) / \partial \boldsymbol{\eta}}{P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} \right)' \right\}. \end{aligned} \quad (\text{A.30})$$

Os elementos da segunda parcela de (A.30) são obtidos por (2.18) e (3.3), respectivamente. Com relação à primeira parcela, de (3.4) temos que

$$\begin{aligned} & \frac{\partial^2 P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'_k} = \\ & = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{kjit} - P_{kit}) \frac{W_{kit}}{P_{kit}^* Q_{kit}^*} \left(\frac{\partial P_{kit}}{\partial \boldsymbol{\zeta}_i} \right) \right\} \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right)' P(\mathbf{U}_{kj..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\eta}_k) d\boldsymbol{\theta}. \end{aligned}$$

Segue que a primeira parcela de (A.30) pode ser escrita como

$$\frac{\partial^2 P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k) / \partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'_k}{P(\mathbf{U}_{kj..}|\boldsymbol{\zeta}, \boldsymbol{\eta}_k)} = \int_{\mathbb{R}^T} \left\{ \sum_{t \in \tau_i} (U_{kjit} - P_{kit}) \frac{W_{kit}}{P_{kit}^* Q_{kit}^*} \left(\frac{\partial P_{kit}}{\partial \boldsymbol{\zeta}_i} \right) \right\} \left(\frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \right)' g_{kj}^*(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Sejam

$$\mathbf{H}_{ikP}(\boldsymbol{\zeta}_i, \boldsymbol{\eta}) = \frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\eta}'_k}, \quad \mathbf{H}_{*P}(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \begin{pmatrix} \mathbf{H}_{11P}(\boldsymbol{\zeta}_1, \boldsymbol{\eta}_1) & \cdots & \mathbf{H}_{1KP}(\boldsymbol{\zeta}_1, \boldsymbol{\eta}_K) \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{n1P}(\boldsymbol{\zeta}_n, \boldsymbol{\eta}_1) & \cdots & \mathbf{H}_{nKP}(\boldsymbol{\zeta}_n, \boldsymbol{\eta}_K) \end{pmatrix}.$$

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

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

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

$$\widehat{\boldsymbol{\psi}}^{(k+1)} = \widehat{\boldsymbol{\psi}}^{(k)} - [\mathbf{H}_{IP}(\widehat{\boldsymbol{\psi}}^{(k)})]^{-1} \mathbf{f}_{IP}(\widehat{\boldsymbol{\psi}}^{(k)}).$$

A.5.1 Expressões para o Algoritmo EM

Nesta seção obtemos as expressões para aplicação do algoritmo EM. Consideramos fixas as quantidades $f_{kit}(\boldsymbol{\theta})$, $r_{kit}(\boldsymbol{\theta})$ e os parâmetros populacionais. Com a utilização da independência entre os itens, precisaremos apenas das derivadas segundas da log-verossimilhança com relação a $\boldsymbol{\zeta}_i$, $i = 1, \dots, n$. De (4.6) temos que

$$\begin{aligned} \mathbf{H}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_i) &= \sum_{k=1}^K \int_{\mathbb{R}^T} \sum_{t \in \tau_i} W_{kit} \left\{ \left[r_{kit}(\boldsymbol{\theta}) - f_{kit}(\boldsymbol{\theta}) \left(\frac{\partial P_{kit}}{\partial \boldsymbol{\zeta}_i} \right) \right] \mathbf{h}'_{kit} + [r_{kit} - P_{kit} f_{kit}] \right. \\ &\quad \left. \times [\mathbf{H}_{kii} - W_{kit}(1 - 2P_{kit}) \mathbf{h}_{kit} \mathbf{h}'_{kit}] \right\} d\boldsymbol{\theta}. \end{aligned}$$

A.6 Curvas de Crescimento: Caso Geral

Sejam

$$\begin{aligned} \Delta_{ks}^{(2)} &= \frac{\partial \Delta_k^{(1)}}{\partial \eta_{1ks}}, \quad \Delta_k^{(2)} = \begin{pmatrix} \Delta_{k1}^{(2)} \\ \vdots \\ \Delta_{kp_k}^{(2)} \end{pmatrix}, \\ \mathbf{m}_{ks}^{(2)} &= \frac{\partial \mathbf{m}_k^{(1)}}{\partial \eta_{1ks}} \quad \text{e} \quad \mathbf{M}_k^{(2)} = \frac{\partial \mathbf{m}_k^{(1)}}{\partial \boldsymbol{\eta}_{1k}} = \left(\mathbf{m}_{k1}^{(2)}, \mathbf{m}_{k2}^{(2)}, \dots, \mathbf{m}_{kp_k}^{(2)} \right). \end{aligned}$$

Então,

$$\begin{aligned} \frac{\partial \log g(\boldsymbol{\theta} | \boldsymbol{\eta}_k)}{\partial \eta_{1ks} \partial \boldsymbol{\eta}'_{1k}} &= \frac{\partial}{\partial \eta_{1ks}} \left\{ \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \Delta_k^{(1)} - \mathbf{m}_k^{(1)'} \right\} \\ &= \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \Delta_{ks}^{(2)} - \mathbf{m}_{ks}^{(2)'}. \end{aligned}$$

Segue que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{1k} \partial \boldsymbol{\eta}'_{1k}} = (\boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1}) \otimes \boldsymbol{\Delta}_k^{(2)} - \mathbf{M}_k^{(2)},$$

onde \otimes representa o *Produto de Kronecker* definido por

$$(\boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1}) \otimes \boldsymbol{\Delta}_k^{(2)} = \begin{pmatrix} \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_{k1}^{(2)} \\ \vdots \\ \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_{kp_k}^{(2)} \end{pmatrix}.$$

Seja $\boldsymbol{\eta}_{2k} = (\eta_{2k1}, \eta_{2k2}, \dots, \eta_{2kq_k})'$ o vetor de elementos da matriz de covariância, $\boldsymbol{\Sigma}_k$, e $\mathbf{V}_k = \boldsymbol{\Sigma}_k^{-1}$. Também

$$\mathbf{V}_{kr}^{(1)} = \frac{\partial \mathbf{V}_k}{\partial \eta_{2kr}}, \quad \mathbf{V}_k^{(1)} = \begin{pmatrix} \mathbf{V}_{k1}^{(1)} \\ \vdots \\ \mathbf{V}_{kq_k}^{(1)} \end{pmatrix},$$

$$\mathbf{m}_{2kr}^{(2)} = \frac{\partial \mathbf{m}_k^{(1)}}{\partial \eta_{2kr}} \quad \text{e} \quad \mathbf{M}_{2k}^{(2)} = (\mathbf{m}_{2k1}^{(2)}, \mathbf{m}_{2k2}^{(2)}, \dots, \mathbf{m}_{2kq_k}^{(2)}).$$

Segue de (8.7) que

$$\begin{aligned} \frac{\partial \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \eta_{2kr} \partial \boldsymbol{\eta}'_{1k}} &= \frac{\partial}{\partial \eta_{2kr}} \left\{ \boldsymbol{\theta}' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_k^{(1)} - \mathbf{m}_k^{(1)'} \right\} \\ &= \boldsymbol{\theta}' \left(\frac{\partial \boldsymbol{\Sigma}_k^{-1}}{\partial \eta_{2kr}} \right) \boldsymbol{\Delta}_k^{(1)} - \left(\frac{\partial \mathbf{m}_k^{(1)'}}{\partial \eta_{2kr}} \right) \\ &= \boldsymbol{\theta}' \mathbf{V}_{kr}^{(1)} \boldsymbol{\Delta}_k^{(1)} - \mathbf{m}_{2kr}^{(2)'}. \end{aligned}$$

Com isso, temos que

$$\frac{\partial^2 \log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_{2k} \partial \boldsymbol{\eta}'_{1k}} = \boldsymbol{\theta}' \otimes \mathbf{V}_k^{(1)} \otimes \boldsymbol{\Delta}_k^{(1)} - \mathbf{M}_{2k}^{(2)'},$$

onde

$$\boldsymbol{\theta}' \otimes \mathbf{V}_k^{(1)} \otimes \boldsymbol{\Delta}_k^{(1)} = \begin{pmatrix} \boldsymbol{\theta}' \mathbf{V}_{k1}^{(1)} \boldsymbol{\Delta}_k^{(1)} \\ \vdots \\ \boldsymbol{\theta}' \mathbf{V}_{kq_k}^{(1)} \boldsymbol{\Delta}_k^{(1)} \end{pmatrix}.$$

As expressões das derivadas segundas de $\log g(\boldsymbol{\theta}|\boldsymbol{\eta}_k)$ com relação a $\boldsymbol{\eta}_{2k}$ para as matrizes de covariâncias estudadas neste trabalho são dadas nas Seções 3.3 a 3.7.

A.7 Curvas de Crescimento: Heitjan

Nesta seção são apresentadas algumas expressões necessárias para o processo de estimação em curvas de crescimento, usando o modelo de Heitjan. Por (8.7), serão necessárias as derivadas de μ_{kt} com relação aos parâmetros das curvas de crescimento, $\boldsymbol{\tau}_k = (\tau_{1k}, \tau_{2k}, \tau_{3k}, \tau_{5k})'$. Considerando

$$h_*^{(i)} = \frac{\partial h_*}{\partial \tau_{ik}} \quad \text{e} \quad h_*^{(i,l)} = \frac{\partial^2 h_*}{\partial \tau_{ik} \partial \tau_{lk}}.$$

temos que, para as derivadas primeiras,

$$\begin{aligned} h_{1kt}^{(1)} &= \frac{\partial h_{kt}}{\partial \tau_{1k}} = -\tau_{4k}[1 + h_{1kt}], \\ h_{1kt}^{(2)} &= \frac{\partial h_{kt}}{\partial \tau_{2k}} = \tau_{4k}[1 + h_{1kt}], \\ h_{1kt}^{(3)} &= \frac{\partial h_{kt}}{\partial \tau_{3k}} = 0, \\ h_{1kt}^{(4)} &= \frac{\partial h_{kt}}{\partial \tau_{4k}} = (\tau_{2k} - \tau_{1k})[1 + h_{1kt}]. \end{aligned}$$

$$\begin{aligned} h_{2kt}^{(1)} &= \frac{\partial h_{kt}}{\partial \tau_{1k}} = 0, \\ h_{2kt}^{(2)} &= \frac{\partial h_{kt}}{\partial \tau_{2k}} = \tau_{4k} h_{2kt} \log h_{2kt}, \\ h_{2kt}^{(3)} &= \frac{\partial h_{kt}}{\partial \tau_{3k}} = \tau_{3k}^{-1} h_{2kt} \log h_{2kt}, \\ h_{2kt}^{(4)} &= \frac{\partial h_{kt}}{\partial \tau_{4k}} = \tau_{2k} h_{2kt} \log h_{2kt}. \end{aligned}$$

$$\begin{aligned} h_{kt}^{(1)} &= \frac{\partial h_{kt}}{\partial \tau_{1k}} = -\tau_{4k}[1 + h_{1kt}]h_{2kt}, \\ h_{kt}^{(2)} &= \frac{\partial h_{kt}}{\partial \tau_{2k}} = \tau_{4k} h_{kt} \{\log h_{2kt} + h_{1kt}^{-1} + 1\}, \\ h_{kt}^{(3)} &= \frac{\partial h_{kt}}{\partial \tau_{3k}} = -h_{kt}(t - t_1)e^{\tau_{2k}\tau_{4k}}, \\ h_{kt}^{(4)} &= \frac{\partial h_{kt}}{\partial \tau_{4k}} = \tau_{4k} h_{kt} \log h_{2kt} + (h_{kt} + h_{2kt})(\tau_{2k} - \tau_{1k}). \end{aligned}$$

$$\begin{aligned}
h_{3kt}^{(1)} &= \frac{\partial h_{3kt}}{\partial \tau_{1k}} = h_{3kt} e^{-\tau_{3k}(t-t_1)}, \\
h_{3kt}^{(2)} &= \frac{\partial h_{3kt}}{\partial \tau_{2k}} = -h_{3kt} e^{-\tau_{3k}(t-t_1)}, \\
h_{3kt}^{(3)} &= \frac{\partial h_{3kt}}{\partial \tau_{3k}} = -(t-t_1) h_{3kt} \log h_{3kt}, \\
h_{3kt}^{(4)} &= \frac{\partial h_{3kt}}{\partial \tau_{4k}} = 0.
\end{aligned}$$

e, para as derivadas segundas,

$$\begin{aligned}
h_{kt}^{(1,1)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{1k}^2} = -\tau_{4k} h_{kt}^{(1)}, \\
h_{kt}^{(1,2)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{1k} \partial \tau_{2k}} = -\tau_{4k} \{h_{kt}^{(2)} + h_{2kt}^{(2)}\}, \\
h_{kt}^{(1,3)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{1k} \partial \tau_{3k}} = -\tau_{4k} \{h_{kt}^{(3)} + h_{2kt}^{(3)}\}, \\
h_{kt}^{(1,4)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{1k} \partial \tau_{4k}} = \tau_{4k} h_{kt} \log h_{2kt} + (h_{kt} + h_{2kt})(\tau_{2k} - \tau_{1k}), \\
h_{kt}^{(2,2)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{1k}^2} = h_{kt}^{-1} [h_{kt}^{(2)}]^2 + \tau_{4k} \{h_{2kt}^{(2)} - h_{2kt} h_{1kt}^{-2} h_{1kt}^{(2)}\}, \\
h_{kt}^{(2,3)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{2k} \partial \tau_{3k}} = h_{kt}^{-1} h_{kt}^{(2)} h_{kt}^{(3)} + \tau_{4k} \{h_{2kt}^{(3)} h_{2kt}^{-1} - h_{1kt}^{-2} h_{1kt}^{(3)}\}, \\
h_{kt}^{(2,4)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{2k} \partial \tau_{4k}} = \{h_{kt} + \tau_{4k} h_{kt}^{(4)}\} [\tau_{4k} h_{kt}]^{-1} h_{kt}^{(2)} + \tau_{4k} h_{kt} \{h_{2kt}^{(4)} h_{2kt}^{-1} - h_{1kt}^{-2} h_{1kt}^{(4)}\}, \\
h_{kt}^{(3,3)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{3k}^2} = h_{kt}^{-1} [h_{kt}^{(3)}]^2, \\
h_{kt}^{(3,4)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{3k} \partial \tau_{4k}} = h_{kt}^{(3)} \{h_{kt}^{(4)} h_{kt}^{-1} + \tau_{2k}\}, \\
h_{kt}^{(4,4)} &= \frac{\partial^2 h_{kt}}{\partial \tau_{4k}^2} = \tau_{2k} [h_{kt}^{(4)} \log h_{2kt} + h_{2kt} h_{kt}^{(4)}] + (h_{kt}^{(4)} + h_{2kt}^{(4)})(\tau_{2k} - \tau_{1k}).
\end{aligned}$$

$$\begin{aligned}
h_{3kt}^{(1,1)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{1k}^2} = h_{3kt}^{(1)} e^{-\tau_{3k}(t-t_1)}, \\
h_{3kt}^{(1,2)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{1k} \partial \tau_{2k}} = -h_{3kt}^{(2)} e^{-\tau_{3k}(t-t_1)}, \\
h_{3kt}^{(1,3)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{1k} \partial \tau_{3k}} = h_{3kt}^{(3)} [e^{-\tau_{3k}(t-t_1)} - (t-t_1)], \\
h_{3kt}^{(2,2)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{1k}^2} = -h_{3kt}^{(2)} e^{-\tau_{3k}(t-t_1)} \\
h_{3kt}^{(2,3)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{2k} \partial \tau_{3k}} = -h_{3kt}^{(3)} [e^{-\tau_{3k}(t-t_1)} - (t-t_1)], \\
h_{3kt}^{(3,3)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{3k}^2} = -(t-t_1) h_{3kt}^{(3)} [\log h_{3kt}^{(3)} + 1], \\
h_{3kt}^{(i,4)} &= \frac{\partial^2 h_{3kt}}{\partial \tau_{ik} \partial \tau_{4k}} = 0, \quad i = 1, \dots, 4.
\end{aligned}$$

Com essas expressões podemos obter as derivadas segundas de μ_{kt} com relação a τ_k . Considerando

$$\mu_{kt}^{(i,l)} = \frac{\partial^2 \mu_{kt}}{\partial \tau_i \partial \tau_l},$$

temos que

Caso $\tau_{4k} \neq 0$:

$$\begin{aligned}
\mu_{kt}^{(1,1)} &= \tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 2} \{ (\tau_{4k}^{-1} + 1) [h_{kt}^{(1)}]^2 - [1 + h_{kt}] h_{kt}^{(1,1)} \} - e^{\tau_{11}} 1_{\{k=1\}}, \\
\mu_{kt}^{(1,2)} &= -\tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \{ h_{kt}^{(1)} + h_{kt}^{(1,2)} - (\tau_{4k}^{-1} + 1) [1 + h_{kt}]^{-1} h_{kt}^{(1)} h_{kt}^{(2)} \}, \\
\mu_{kt}^{(1,3)} &= -\tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \{ h_{kt}^{(1,3)} - (\tau_{4k}^{-1} + 1) [1 + h_{kt}]^{-1} h_{kt}^{(1)} h_{kt}^{(3)} \}, \\
\mu_{kt}^{(1,4)} &= [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \left\{ \tau_{4k}^{-2} e^{\tau_{2k}} h_{kt}^{(1)} - \tau_{4k}^{-1} e^{\tau_{2k}} \left[h_{kt}^{(1)} \left(-(\tau_{4k}^{-1} + 1) [1 + h_{kt}]^{-1} h_{kt}^{(4)} + \right. \right. \right. \\
&\quad \left. \left. \left. + \tau_{4k}^{-2} \log(1 + h_{kt}) \right) + h_{kt}^{(1,4)} \right] \right\}, \\
\mu_{kt}^{(2,2)} &= \mu_{kt}^{(2)} - \tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \{ h_{kt}^{(2)} + h_{kt}^{(2,2)} - (\tau_{4k}^{-1} + 1) [1 + h_{kt}]^{-1} [h_{kt}^{(2)}]^2 \}, \\
\mu_{kt}^{(2,3)} &= -\tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \{ h_{kt}^{(2,3)} - (\tau_{4k}^{-1} + 1) [1 + h_{kt}]^{-1} h_{kt}^{(2)} h_{kt}^{(3)} \}, \\
\mu_{kt}^{(2,4)} &= \mu_{kt}^{(4)} + \tau_{4k}^{-2} e^{-\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \{ \tau_{4k}^{-1} h_{kt}^{(2)} \log(1 + h_{kt}) + h_{kt}^{(2,2)} - \\
&\quad - 2[1 + h_{kt}]^{-1} h_{kt}^{(2)} h_{kt}^{(4)} + h_{kt}^{(2)} - h_{kt}^{(2,4)} \}, \\
\mu_{kt}^{(3,3)} &= -\tau_{4k}^{-1} e^{\tau_{2k}} [1 + h_{kt}]^{-\tau_{4k}^{-1} - 1} \{ h_{kt}^{(3,3)} - (\tau_{4k}^{-1} + 1) [1 + h_{kt}]^{-1} [h_{kt}^{(3)}]^2 \}, \\
\mu_{kt}^{(3,4)} &= -[\tau_{4k}^{-1} h_{kt}^{(3)} [1 + h_{kt}]^{-1}] \mu_{kt}^{(4)} + \tau_{4k}^{-2} e^{-\tau_{2k}} h_{kt}^{(3)} [1 + h_{kt}]^{-2} \{ 1 + \tau_{4k}^{-1} [1 + h_{kt}]^{-1} h_{kt}^{(4)} \}, \\
\mu_{kt}^{(4,4)} &= -\tau_{4k}^{-1} \mu_{kt}^{(4)} [1 + [1 + h_{kt}]^{-1} h_{kt}^{(4)}] - \tau_{4k}^{-1} e^{-\tau_{2k}} [1 + h_{kt}]^{-2} \{ [h_{kt}^{(4,4)} + \\
&\quad + (1 + h_{kt})^{-1}]^{-1} [h_{kt}^{(4)}]^2 \}.
\end{aligned}$$

Caso $\tau_{4k} = 0$:

$$\begin{aligned}
\mu_{kt}^{(1,1)} &= e^{\tau_{2k}} h_{3kt}^{(1,1)} - e^{\tau_{11}} 1_{\{k=1\}}, \\
\mu_{kt}^{(1,2)} &= e^{\tau_{2k}} [h_{3kt}^{(1)} + h_{3kt}^{(1,2)}], \\
\mu_{kt}^{(1,3)} &= e^{\tau_{2k}} h_{3kt}^{(1,3)}, \\
\mu_{kt}^{(2,2)} &= \mu_{kt}^{(2)} + e^{\tau_{2k}} [h_{3kt}^{(1)} + h_{3kt}^{(2,2)}], \\
\mu_{kt}^{(2,3)} &= \mu_{kt}^{(3)} + e^{\tau_{2k}} h_{3kt}^{(2,3)}, \\
\mu_{kt}^{(3,3)} &= e^{\tau_{2k}} h_{3kt}^{(3,3)}.
\end{aligned}$$

Referências Bibliográficas

- [1] Ackerman, T. (1996). Developments in Multidimensional Item Response Theory. *Applied Psychological Measurement*. Special Issue: Vol. **20**, Dec. 1996, Number 4.
- [2] Andersen E. B. (1980). *Discrete Statistical Models with Social Science Applications*. New York: North-Holland Publishing Company.
- [3] Andrade, D. F. (1999). *Comparando o Desempenho de Grupos (Populações) de Respondentes Através da Teoria da Resposta ao Item*. Tese apresentada ao Departamento de Estatística e Matemática Aplicada da UFC para o concurso de professor titular.
- [4] Andrade, D. F. e Klein, R. (1999). Métodos estatísticos para avaliação educacional : teoria da resposta ao item. *Boletim da ABE*, *43*, 21-28.
- [5] Andrade, D. F. , Tavares, H. R., Valle, R. C. (2000). *Introdução À Teoria da Resposta ao Item: Conceitos e Aplicações*. 14º SINAPE: Caxambu, MG.
- [6] Andrade, D. F. e Valle, R. C. (1998). Introdução à teoria da resposta ao item: conceitos e aplicações. *Estudos em Avaliação Educacional*, *18*, 13-32.
- [7] Baker, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- [8] Batista, J.R. (1999). *Valores Plausíveis para Estimação de Parâmetros Populacionais em Modelos da Teoria da Resposta ao Item*. Dissertação de Mestrado. Belo Horizonte: ICEX/UFMG.
- [9] Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New york: John Wiley and Sons.
- [10] Birnbaum, A. (1957). *Efficient design and use of tests of a mental ability for various decision-making problems*, (Series Report No. 58-16. Project No. 7755-23). USAF School of Aviation Medicine, Texas: Randolph Air Force Base.
- [11] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M. Lord and M.R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA:Addison-Wesley.

-
- [12] Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm. *Psychometrika*, 46, 433-459.
- [13] Bock, R. D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- [14] Bock, R. D. and Zimowski, M. F. (1997). Multiple Group IRT. In W.J. van der Linden e R.K. Hambleton Eds. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- [15] Chow, Y.S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.
- [16] Cochran, W. G., Cox, G. M. (1950). *Experimental Designs*. New York: Wiley
- [17] Dempster, A. P. , Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- [18] Doornik, J.A. (2000). *Object-Oriented Matrix Programming using Ox 2.0*. London: Timberlake Consultants Ltd and Oxford: www.nuff.ox.ac.uk/Users/Doornik.
- [19] Esposito, Y. L. et al (1995). *Avaliação do processo de inovações no ciclo básico e seu impacto sobre a situação de ensino-aprendizagem na região metropolitana de São Paulo*. São Paulo: Fundação Carlos Chagas/Departamento de Pesquisas Educacionais, 3 v.
- [20] Genz, A. C. and Malik, A. A. (1980). An adaptive algorithm for numerical integration over a N -retangular region. *J. Comput. Appl. Math.*, 6, 295-302.
- [21] Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics*. Belmont, CA: Wadsworth Publishing Company, Inc.
- [22] Haberman, S. (1975). *Maximum Likelihood Estimates in Exponential Response Models*, (Technical Report) Chicago, IL: University of Chicago.
- [23] Haley, D. C. (1952). *Estimation of the dosage mortality relationship when the dose is subject to error*, (Technical Report, 15) Stanford, Calif.: Stanford University, Applied Mathematics and Statistics Laboratory.
- [24] Hambleton, R. K. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Academic Publishers.
- [25] Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications.

-
- [26] Hedges, L. V. and Vevea, J. L. (1997). A study of equating in NAEP. *Paper presented at The NAEP Validity Studies Panel*. Palo Alto : American Institutes for Research.
- [27] Heitjan, D. F. (1991a). Generalized Norton-Simon models of tumour growth. *Statistics in Medicine*, 10. 1075-1088.
- [28] Heitjan, D. F. (1991b). Nonlinear modeling of serial immunologic data: A case study. *Journal of the American Statistical Association*, 86. 891-898.
- [29] Issac, E. and Keller, H. B. (1966). *Analysis of Numerical Methods*. New York: Wiley & Sons.
- [30] Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*. Chapman and Hall: London.
- [31] Kolen, M. J. and Brennan, R. L. (1995). *Test Equating - Methods and Practices*. New York: Springer.
- [32] Lindsey, J.K. (1993). *Models for Repeated Measurements*. Oxford: Clarendon Press.
- [33] Lord, F. M. (1952). A theory of test scores (No. 7). *Psychometric Monograph*.
- [34] Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.
- [35] Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39, 247-264.
- [36] Lord, F. M. (1975). *Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters*, (Research Bulletin RB-75-33). Princeton, NJ: Educational Testing Service.
- [37] Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- [38] Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, 49, 359-381.
- [39] Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177-195.
- [40] Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177-196.
- [41] Mislevy, R. J. and Stocking, M. L. (1989). A Consumer's Guide to LOGISTIC and BILOG. *Applied Psychological Measurement*, 13 57-75.

-
- [42] Mislevy, R. J. and Bock, R. D. (1990). *BILOG 3 : Item Analysis and Test Scoring with Binary Logistic Models*. Chicago : Scientific Software, Inc.
- [43] Nelder, J. A. (1961). The fitting of a generalization of the logistic curve. *Biometrika*, 17 , 89-100.
- [44] Nelder, J. A. (1962). An alternative form of a generalized logistic equation. *Biometrics*, 18 , 614-616.
- [45] Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, 16 (1), 1-32.
- [46] Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3,1-18.
- [47] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley & Sons.
- [48] Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W.J. van der Linder e R.K. Hambleton Eds. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- [49] Sandland, R. L. and McGilchrist, C. A. (1979). Stochastic grow curve analysis. *Biometrics*, 35, 255-271.
- [50] Sen, P. K., Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction With Applications*. New York: Chapman & Hall.
- [51] Singer, J. M., Andrade, D. A. (2000). Analysis of longitudinal data. In P. K. Sen and C. R. Rao, Eds. *Handbook of Statistics*, 18 (p. 115-160).
- [52] Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- [53] Stroud, A. H. (1971). *Aproximate Calculation os Multiple Integrals*. New Jersey: Prentice Hall, Englewood Cliffs.
- [54] Stroud, A. H. and Secrest, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, New Jersey : Prentice-Hall.
- [55] Swaminathan, H. and Gifford J. A. (1983). Estimation of Parameters in the Three-Parameter Latent Trait Model. In D. Weiss (Ed.), *New Horizons in Testing*. New York: Academic Press.
- [56] Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, 11, 1-13.

- [57] Valle, R. C. (1999). *Teoria da Resposta ao Item..* Dissertação de Mestrado. São Paulo: IME/USP.
- [58] Van der Linden, W.J. and Hambleton, R.K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- [59] Van Dooren, P. and De Ridder, L. (1976). An adaptive algorithm for numerical integration over a N -retangular cube. *J. Comput. Appl. Math.*, 2, 207-217.
- [60] Wingersky, M.S., Barton, M.A. and Lord, F.M. (1982). *LOGIST User's Guide*. Princeton, NJ: Educational Testing Service.
- [61] Wood, R.L., Wingersky, M.S., Lord, F.M. (1976). *LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters* (RM-76-6). Princeton, NJ: Educational Testing Service.
- [62] Zimowski, M. F. , Muraki, E. , Mislevy, R. J. and Bock, R.D. (1996). *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Chicago : Scientific Software, Inc.