UNIVERSIDADE FEDERAL DO PARÁ
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

PAULO FERNANDO RODRIGUES

ESTIMAÇÃO BAYESIANA PARA CURVAS DE CRESCIMENTO EM MODELOS DE RESPOSTA AO ITEM

Belém/PA
2006

PAULO FERNANDO RODRIGUES

ESTIMAÇÃO BAYESIANA PARA CURVAS DE CRESCIMENTO EM MODELOS DE RESPOSTA AO ITEM

Dissertação apresentada ao Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará como requisito para obtenção do Título de Mestre em Matemática e Estatística.

Orientadora: Profa. Dra. Maria Regina Madruga Tavares

Co-Orientador: Prof. Dr. Héliton Ribeiro Tavares

Belém/PA
2006

PAULO FERNANDO RODRIGUES

ESTIMAÇÃO BAYESIANA PARA CURVAS DE CRESCIMENTO EM MODELOS DE RESPOSTA AO ITEM

Banca Examinadora

Profa. Dra. Maria Regina Madruga Tavares
Universidade Federal do Pará
Orientadora

Prof. Dr. Marcus Pinto da Costa Rocha
Universidade Federal do Pará
Examinador

Prof. Dra. Alcione Miranda dos Santos
Universidade Federal do Maranhão
Examinador

DEDICATÓRIA

Aos meus pais e avó...
Pessoas especiais
que desde cedo me ensinaram
o valor do saber.

AGRADECIMENTOS

Primeiramente a **Deus**, pois sem ele nada seria possível.

À minha orientadora Profa. Dra. Maria Regina Madruga Tavares e co-orientador Prof. Heliton Ribeiro Tavares, por sempre terem acreditado na realização desta dissertação.

A todos os professores integrantes do Programa de Pós-Graduação em Matemática e Estatística, pela grande contribuição ao meu aprendizado ao longo de todo este curso.

À minha esposa Márcia, meus filhos Fernanda e Fernando, pela paciência e comprensão, pelo pouco tempo disponível nestes dois anos.

Em fim, a todos que contribuíram de alguma maneira para a complementação deste trabalho.

Resumo

RODRIGUES, Paulo Fernando. Estimação Bayesiana para Curvas de Crescimento em Modelos de Resposta ao Item. 2006. Dissertação (Mestrado em Matemática e Estatística)-PPGME/UFPA, Belém - PA, Brasil.

Neste trabalho, considera-se a situação em que indivíduos são submetidos ao longo do tempo a instrumentos de avaliação (testes), com o objetivo de medir seus desempenhos (ou habilidades) em alguma área de interesse. Por conta disto, é natural admitir-se a existência de uma estrutura de covariância entre esses desempenhos (habilidades/proficiências) no decorrer do tempo. Os desempenhos são obtidos a partir de respostas dadas aos testes, formados por *itens* (questões) com características conhecidas, e utilizando-se Modelos da Teoria de Resposta ao Item (TRI). Para modelar a dependência entre as habilidades nos vários períodos observados, são estudadas duas estruturas de covariância: a *Matriz de Covariância Uniforme* e a *Matriz de Covariância não Estruturada* (ou *geral*). O crescimento das habilidades adquiridas ao longo do tempo é modelado por curvas de crescimento (lineares e quadráticas), e propõe-se a metodologia Bayesiana para a estimação dos parâmetros de dispersão da distribuição das habilidades e dos parâmetros da curva de crescimento. São apresentados resultados empíricos, obtidos através da realização de simulações, e resultados de uma aplicação a dados reais, obtidos do Projeto de Desenvolvimento da Escola realizado pelo Governo Federal no período de 1999 a 2003.

Palavras-chave: Teoria da Resposta ao Item, Metodologia Bayesiana, Estrutura de Covariância, Curvas de Crescimento, Habilidades.

Abstract

RODRIGUES, Paulo Fernando. Bayesian Estimation for Growth Curves in Item Response Models. 2006. Thesis (Master in Matematic and Statistic)- PPGME/UFPA, Belém - PA, Brazil.

In this work, is considered the situation in that individuals they are submitted along the time to evaluation instruments (tests), with the objective of measuring your performance (or abilities) in some area of interest. Due to this, it is natural to admit the existence of a covariance structure among those performance (abilities/proficiencies) in elapsing of the time. The performances are obtained starting from answers given to the tests, formed by items (subjects) with characteristics known, and being used Models of the Item Response Theory (IRT). To model the dependence among the abilities in the several ones observed periods, are studied two covariance structures: the Uniform Covariance matrix and the Unstructured Covariance matrix (or general). The growth of the acquired abilities along the time are modeled by growth curves (lineal and quadratic), and intends the Bayesiana methodology for the estimate of the abilities distribution parameters of dispersion and of the growth curve parameters. Empiric results are presented, obtained through the accomplishment of simulations, and results of an application to real data, obtained of the Project of Development of the School accomplished by the Federal Government in the period from 1999 to 2003.

Keywords: Item Response Theory, Bayesiana Methodology, Covariance Structures, Growth Curves, Abilities.

Índice

Lista de Tabelas

Capítulo 1

Introdução

Em avaliações psicológicas ou educacionais a variável de interesse, na maioria das vezes, não é diretamente observável no indivíduo. Neste caso, ela é dita ser uma variável não observável, habilidade, proficiência ou traço latente. Embora essas variáveis sejam características implícitas a cada ser humano, elas podem ser facilmente descritas e listadas, como, por exemplo, a inteligência, a habilidade em executar uma tarefa, o grau de ansiedade, o nível de entendimento de um texto, etc. Elas não podem ser medidas diretamente como o peso ou a altura de uma pessoa, apesar de todas serem características implícitas a cada ser humano.

O objetivo das avaliações educacionais e psicológicas é medir traços latentes no indivíduo examinado a partir da observação de variáveis secundárias, tipo acerto/erro, que estejam relacionadas a eles através de algum instrumento de medida. É razoável admitir a hipótese de que cada respondente responda a um item de acordo com habilidades implícitas. Esses itens podem ser dicotômicos (certo/errado) ou não, e o interesse pode estar em medir habilidades em várias populações de indivíduos.

Admitiremos no nosso estudo que se deseje medir apenas uma habilidade, a qual representaremos pela letra grega θ. No caso de itens dicotômicos, existirá uma probabilidade que o respondente j da população k, com habilidade θ_{jk}, dará uma resposta correta ao item i, representada por P_{jik}. No caso típico de testes com itens dicotômicos a probabilidade será pequena se a habilidade do respondente for pequena ou será grande se a mesma também o for.

A TRI (Teoria da Resposta ao Item), baseia-se em um conjunto de modelos matemáticos que buscam representar a relação entre a probabilidade de um indivíduo dar uma determinada resposta a um item, como função de algumas características do item e da habilidade (ou habilidades) do respondente [ver Andrade et al.(2000)]. No caso dicotômico aplicado à área educacional, esta relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto ao item.

Segundo Baker (1992), muitos autores ajudaram na construção da teoria de resposta ao item (TRI), em especial dois desses autores, para os quais é importante citar seus nomes e respectivos trabalhos, Lord e Rash. O trabalho de Lord (1952) ganhou força nos anos 50 com a aplicação em testes educacionais. Lord iniciou o desenvolvimento formal da teoria, e também contribuiu para o desenvolvimento de programas para computadores necessários para pôr a

teoria em prática, o que levou posteriormente a redigir um livro clássico junto com M. Novick (1968). O matemático dinamarquês Rasch (1960) reforçou os estudos da TRI criando o modelo de 1 parâmetro.

Nas últimas décadas, a TRI vem tornando-se a técnica predominante no contexto de avaliações educacionais em vários países. No Brasil, a primeira experiência ocorreu em 1995, na análise de dados do SAEB (Sistema Nacional de Avaliação da Educação Básica). A TRI permite uma melhor análise de cada item que constitui o instrumento de avaliação e/ou medida, levando em consideração suas características na estimação das habilidades. Além disso, a TRI é extremamente vantajosa, pois permite-nos comparar os resultados produzidos por grupos de indivíduos diferentes, mesmo quando são aplicados testes parcialmente ou até totalmente diferentes, o que não ocorre com a teoria clássica. Essa propriedade é útil, particularmente, quando se deseja avaliar a evolução da medida de uma habilidade ao longo do tempo, o que se denomina de estudos longitudinais (Singer and Andrade (2000)).

Dois pontos básicos para o desenvolvimento de modelos da TRI são: *i*) a performance de um respondente em um teste pode ser prevista por sua habilidade ou traço latente; *ii*) a relação entre a habilidade do indivíduo e a sua probabilidade de resposta ao item pode ser descrita por uma função, conhecida como curva característica do item (CCI), parametrizada pelas características do item de um teste, também chamadas de *parâmetros do item*: discriminação, dificuldade e acerto casual.

Antes de falarmos sobre os modelos da TRI, faz-se necessário uma explicação sobre duas hipóteses básicas na TRI: a *unidimensionalidade* e a *independência local*. Sob a hipótese de unidimensionalidade, admite-se que somente uma habilidade está sendo medida pelo modelo. Isto é, o conjunto de itens deve estar medindo um único traço latente. A suposição de independência local postula que, para uma dada habilidade, as respostas atribuídas aos diferentes itens são independentes entre si, ou seja, o indívíduo não aprende com os itens. Tal pressuposto será importante para a estimação dos parâmetros nos modelos.

Teoricamente, pode existir uma infinidade de modelos da TRI. Porém, poucos modelos são usados na prática, como por exemplo, o modelo logístico unidimensional de 1, 2 ou 3 parâmetros para dados dicotômicos, e os modelos de resposta nominal e de resposta gradual para dados não dicotômicos (veja Andrade et. al. (2000) para detalhes). Os modelos propostos na literatura dependem fundamentalmente de três fatores:

- da natureza dos itens (dicotômicos ou politômicos);
- do número de populações envolvidas;
- da quantidade de traços latentes que está sendo medida.

Em várias situações práticas, diferentes grupos de pessoas são avaliados e os parâmetros das distribuições das habilidades estimados. Um exemplo é o Exame Nacional da Avaliação da Educação Básica (SAEB) onde amostras de estudantes da 4ª e 8ª séries do ensino fundamental e 3ª série do ensino médio são avaliadas a cada dois anos. Um dos principais objetivos é avaliar o possível crescimento dos parâmetros da proficiência (habilidade) média através das séries e anos. Já em outras situações, principalmente no campo educacional, o interesse pode estar na construção de escalas de proficiência. Por exemplo, alguém pode construir uma escala de proficiência que compreenda estudantes da 1ª série do ensino fundamental à 3ª série do ensino médio. Para cada um desses 11 grupos a média e a dispersão das habilidades da população precisam ser estimadas. Se pudermos representar o crescimento por uma função com poucos parâmetros, como, por exemplo, uma polinomial de grau 2, o número de parâmetros para representar a habilidade média cairá de 11 para 3, diminuindo substancialmente a demanda computacional e produzindo melhores resultados.

Este trabalho baseia-se na situação onde o mesmo indivíduo é avaliado ao longo do tempo e, portanto, é natural supor a existência de uma estrutura de covariância não trivial (diagonal) entre as habilidades ao longo do tempo. Daí são propostas curvas de crescimento de modelos da TRI para avaliar um grupo de indivíduos em um certo número de instantes. A cada instante, os indivíduos são submetidos a um teste com certa quantidade de itens, tendo ou não itens comuns. Será assumido que todos os parâmetros dos itens envolvidos neste estudo são conhecidos, ou seja, os testes são construídos com itens calibrados na mesma métrica.

Devido à restrição de positividade dos parâmetros de dispersão, ocorrem problemas na convergência do método de Newton-Raphson para obter os estimadores desses parâmetros através do método da máxima verossimilhança. Por conta disto, optamos por utilizar a estimação bayesiana para contornar esses problemas.

O objetivo dessa dissertação é, considerando situações envolvendo dados longitudinais, ou seja, aquelas onde um grupo de indivíduos é acompanhado durante certo período de tempo com alguma estrutura de covariância, estimar os parâmetros da curva de crescimento e os parâmetros de dispersão da distribuição das habilidades utilizando estimadores bayesianos.

A estimação dos parâmetros populacionais já foi objetivo de vários trabalhos. Anderson and Madsen (1977) mostraram que a estimação dos parâmetros da população pode ser realizada diretamente, sem a prévia estimação das habilidades individuais. Neste trabalho ele considerou que os parâmetros dos itens eram conhecidos. Alguns trabalhos posteriores [veja Sanathanan & Blumenthal (1978) e Leeuw & Verhelst (1986)] extenderam esses resultados para a estimação conjunta dos parâmetros dos itens e da distribuição das habilidades, mas somente para o modelo logístico com um parâmetro e no caso de uma única população. Mislevy (1986)

trabalhou com a estimação dos parâmetros da distribuição das habilidades baseado no método da *Máxima Verossimilhança Marginal*. Bock & Zimowski (1997) extenderam os casos anteriores para situações com várias populações independentes. Andrade e Tavares (2000) trabalharam com a estimação dos parâmetros populacionais, considerando dados longitudinais.

Esta Dissertação é composta de cinco capítulos, além deste que é introdutório, onde os conteúdos destes capítulos são resumidos abaixo.

- No Capítulo 2, intitulado Curvas de Crescimento, primeiro fazemos as definições e notações necessárias, apresentamos o modelo a ser utilizado e por fim as diversas estruturas de covariâncias existentes.

- No Capítulo 3, intitulado Inferência Bayesiana, primeiro é feita um pequena introdução à inferência bayesiana, logo após determinamos a função de verossimilhança e as prioris para cada um dos tipos de estruturas de covariância e finalmente apresentamos os estimadores dos parâmetros da curva das crescimento e dos parâmetros de dispersão.

- No Capítulo 4, apresentamos os resultados da teoria desenvolvida, aplicados tanto a dados simulados como a dados reais.

- No Capítulo 5, apresentamos as conclusões e algumas propostas para pesquisas futuras.

Capítulo 2

Curvas de Crescimento

2.1 Definições e notações

Considere que um grupo de indivíduos é avaliado em uma certa área do conhecimento em m instantes do tempo. A cada instante t, os indivíduos são submetidos a um teste com n_t itens cada, tendo ou não itens comuns. O número total de itens distintos será portanto, $n \leq n_c = \sum_{t=1}^{m} n_t$. Consideraremos também que os instantes envolvidos na análise correspondem a série $\mathbf{t} = (t_1, t_2, \cdots, t_m)'$. Se as séries forem sequenciais e os instantes de avaliação equipaçados, será adotado $t_1 = 1,\ t_2 = 2, \cdots,\ t_m = m$.

Seja μ_t a habilidade média do grupo no instante $t, t = 1, ..., m$. Em geral, toma-se $\mu_t = f(t_t|\boldsymbol{\alpha})$ sendo uma função contínua duplamente diferenciável com p parâmetros e $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$' o vetor de parâmetros da função. Daí, o vetor de médias para os m instantes, poderá ser escrito como $\boldsymbol{\mu} = (f(t_1|\boldsymbol{\alpha}), f(t_2|\boldsymbol{\alpha}), \ldots, f(t_m|\boldsymbol{\alpha}))$'.

Muitas funções tem sido propostas na teoria das curvas de crescimento para dados longitudinais. Frequentemente, funções simples, tais como a função linear ou quadrática, são apropriadas para modelar as estruturas longitudinais, mas funções polinomiais de ordem maior que dois, e algumas funções não-lineares, tem sido também aplicadas em vários casos(ver Lindsey (1993) e Singer and Andrade (2000), por exemplo). A Tabela 2.1 apresenta alguns exemplos de funções usadas em várias situações.

Tabela 2.1 *Modelos para curvas de crescimento*

Logístic Tripla	$\mu_t = \frac{\gamma_1}{1+e^{-\alpha_1(t-\beta_1)}} + \frac{\gamma_2}{1+e^{-\alpha_2(t-\beta_2)}} + \frac{\gamma_3}{1+e^{-\alpha_3(t-\beta_3)}}$
Gompertz	$\mu_t = \gamma_1 + \gamma_2 e^{-e^{-\alpha(t-\beta)}}$
Jenns	$\mu_t = \alpha + \beta t - e^{\gamma+\delta t}$
Count	$\mu_t = \alpha + \beta t + \gamma \log t$
Mitscherlish	$\mu_t = \alpha - \beta e^{-\gamma t}$
Polinomial	$\mu_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_p t^p$

Neste trabalho serão consideradas apenas dois tipos de curva de crescimento: polinomial linear e polinomial quadrática.

2.2 Apresentação do Modelo

Seja

$$P_{jit} = P(U_{jit} = 1|\theta_{jt}, \boldsymbol{\zeta}_i) \tag{2.1}$$

uma função de resposta ao item duplamente diferenciável que representa a probabilidade condicional de uma resposta correta ao item $i, i = 1, ..., n_t$, pelo indivíduo $j, j = 1, ..., N$, no teste $t, t = 1, ..., m$, onde U_{jit} representa a resposta (binária), θ_{jt} a habilidade (característica latente) e $\boldsymbol{\zeta}_i$ o vetor de parâmetros (conhecidos) do item. Neste trabalho adotou-se o modelo logístico de 3 parâmetros (ML3) em sua versão unidimensional, cuja forma é

$$P_{jit} = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \qquad \boldsymbol{\zeta}_i = (a_i, b_i, c_i), \tag{2.2}$$

onde a_i é o parâmetro de discriminação do item i, b_i é o parâmetro de dificuldade do item i e c_i é o parâmetro de acerto ao acaso do item i. D é um fator de escala, constante e igual a 1 ou 1.7. O valor 1.7 é usado quando queremos que a função logística renda resultados similares aos da função de distribuição acumulada da distribuição normal padrão.

A seguir, apresenta-se a denominada curva característica de um item (CCI), isto é, a representação sob forma de gráfico de um particular modelo da TRI, enfatizando as propriedades de seus parâmetros.

Como se pode notar, o parâmetro b_i representa o ponto na escala de habilidade em que um respondente tem probabilidade $(1 + c_i)/2$ de responder ao item i corretamente. Portanto, quanto maior o valor de b_i, mais difícil é o item i, e vice-versa.

O parâmetro c_i representa a probabilidade de um respondente com baixa habilidade responder corretamente o item i. Então, quando não for permitido "chutar", teremos $c_i = 0$, como consequência b_i representará o ponto na escala de habilidade onde a probabilidade de acertar o item i é 0,5.

O parâmetro a_i é proporcional à inclinação da curva em torno de b_i. Portanto, itens com a_i negativo não são esperados sob esse modelo. Valores baixos de a_i indicam que o item i tem pouco poder de discriminação (respondentes com habilidades bem diferentes têm aproximadamente a mesma probabilidade de responder corretamente ao item) e valores muito altos indicam itens com curvas características com bastante aclividade, que discriminam os respondentes basicamente em dois grupos: os com habilidades menores que b_i e os com habilidades maiores que b_i.

Assumindo a independência condicional das respostas aos itens no teste t, dado θ_t, temos que

$$P(\boldsymbol{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) = \prod_{i \in I_t} P(U_{jit}|\theta_t, \boldsymbol{\zeta}_i), \tag{2.3}$$

onde $\boldsymbol{I}_t$ é o conjunto de índices dos itens presentes no teste t, $\boldsymbol{U}_{j.t} = (U_{j1t}, ..., U_{jn_tt})'$ é o vetor

(n_t x 1) de respostas do indivíduo j no teste t e $\boldsymbol{\zeta}_t = (\boldsymbol{\zeta}_1^{'}, ..., \boldsymbol{\zeta}_{n_t}^{'})$ o vetor de parâmetros dos itens no teste t. Note que, por conveniência e sem perda de generalidade, o índice j foi retirado do parâmetro de habilidade, pois nosso interesse consiste na distribuição de habilidade a cada instante, t e não em alguma habilidade individual θ_{jt}. Além disso, assumindo independência longitudinal das respostas aos itens ao longo dos m testes, dadas as habilidades nos m testes, temos

$$P(\boldsymbol{U}_{j..}|\boldsymbol{\theta}, \zeta) = \prod_{t=1}^{m} P(\boldsymbol{U}_{j.t}|\theta_t, \boldsymbol{\zeta}) = \prod_{t=1}^{m} \prod_{i \in I_t} P(U_{jit}|\theta_t, \boldsymbol{\zeta}_i) \tag{2.4}$$

com $\boldsymbol{U}_{j..} = (\boldsymbol{U}_{j.1}^{'}, ..., \boldsymbol{U}_{j.m}^{'})'$ sendo o vetor (n_c x 1) de respostas do indivíduo j em todos os testes e $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)'$. Finalmente, assumindo que $\boldsymbol{\theta}$ segue uma distribuição contínua multivariada com vetor de parâmetros $\boldsymbol{\phi}$ de finitos elementos, a probabilidade incondicional, ou marginal, de $\boldsymbol{U}_{j..}$ pode ser escrita como

$$P(\boldsymbol{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\phi}) = \int_{R^m} P(\boldsymbol{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\phi}) d\boldsymbol{\theta}, \tag{2.5}$$

onde $g(\boldsymbol{\theta}|\boldsymbol{\phi})$ é a função densidade da habilidade. A probabilidade acima dependerá dos parâmetros (conhecidos) dos itens e de $\boldsymbol{\theta}$, somente através de sua distribuição, em particular do vetor de parâmetros $\boldsymbol{\phi}$ da distribuição das habilidades, cuja estimação, como vimos, é o principal objetivo deste trabalho.

Neste trabalho, consideramos que a distribuição de $\boldsymbol{\theta}$ é normal multivariada com vetor de parâmetros $\boldsymbol{\phi} = (\boldsymbol{\mu}^{'}, \Psi^{'})'$, onde $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)' = (f(t_1|\boldsymbol{\alpha}), f(t_2|\boldsymbol{\alpha}), \ldots, f(t_m|\boldsymbol{\alpha}))'$ é o vetor de médias nos m instantes, com $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)$' e $\boldsymbol{\Psi} = (\psi_1, ..., \psi_d)'$ é um vetor (d x 1) cujos elementos são escalares desconhecidos que definem $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\Psi})$, a matriz ($m$ x m) de covariâncias, a qual pode assumir diversas formas. Na Seção 2.3 apresentamos algumas estruturas de covariância existentes na literatura.

2.3 Algumas Estruturas de Covariância

Nos estudos longitudinais aparecem diferentes tipos de estruturas de covariância, onde as mais comuns consideradas na literatura são a *uniforme*, *bandas*(1) e $AR(1)$. Um ponto negativo dessas estruturas é que elas assumem que as variâncias são as mesmas ao longo do tempo, algo que não é encontrado em muitos estudos longitudinais. Para concorrer com elas, consideraremos uma estrutura que admite tanto a ocorrência de variâncias diferentes, como de covariâncias iguais, resultando em correlações diferentes, chamada *estrutura heterocedástica*, a qual depende de $m + 1$ parâmetros, e a *não estruturada*, que depende de $m(m + 1)/2$ parâmetros. É também apresentada a estrutura de covariância *diagonal*, que pode ser considerada quando as correlações são pequenas ou nulas.

2.3.1 Estrutura de covariância uniforme

Este tipo de estrutura foi utilizado em vários estudos longitudinais (ver Graybill[1969] para detalhes). Ele assume que todas as variâncias são iguais, assim como todas as covariâncias. Essa estrutura de covariância é representada por

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

onde $\mathbf{\Psi} = (\rho, \sigma)'$.

2.3.2 Estrutura de covariância bandas(1)

Essa estrutura(ver Graybill[1969] e Davinson and Mackinnon[1993] para detalhes), está relacionada ao processo de médias-móveis de ordem 1. A diferença para a estrutura anterior é que somente as covariâncias entre as habilidades em dois instantes consecutivos são diferentes de zero. Essa estrutura de covariância é representada por

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

onde $\mathbf{\Psi} = (\rho, \sigma)'$.

2.3.3 Estrutura de covariância AR(1)

Essa estrutura (ver Davinson and Mackinnon[1993] para detalhes) assume que as correlações entre as habilidades decresce conforme a distância entre os instantes de avaliação cresce. Assim como nas estruturas anteriores, as variâncias são consideradas iguais para todos os instantes. Essa estrutura de covariância é representada por

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

onde $\mathbf{\Psi} = (\rho, \sigma)'$.

2.3.4 Estrutura de covariância heterocedástica

Diferentemente das três estruturas anteriores, nesta estrutura as variâncias variam no decorrer do tempo. Ela assume que as covariâncias são todas iguais, mas não as correlações. Essa estrutura de covariância é representada por

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{12} & \cdots & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 & \sigma_{12} & \cdots & \sigma_{12} \\ \sigma_{12} & \sigma_{12} & \sigma_3^2 & \cdots & \sigma_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{12} & \sigma_{12} & \sigma_{12} & \cdots & \sigma_m^2 \end{pmatrix},$$

onde $\boldsymbol{\Psi} = (\sigma_1^2, \ldots, \sigma_m^2, \sigma_{12})'$.

2.3.5 Estrutura de covariância geral (Não estruturada)

No caso geral, aos elementos da matriz de covariância existe a imposição de serem não estruturados. Ela tem $m(m+1)/2$ parâmetros e é representada por

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2m} \\ \sigma_{13} & \sigma_{12} & \sigma_3^2 & \cdots & \sigma_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \sigma_{3m} & \cdots & \sigma_m^2 \end{pmatrix},$$

onde $\boldsymbol{\Psi} = (\sigma_1^2, \sigma_{12}, \ldots, \sigma_{1m}, \sigma_2^2, \ldots, \sigma_{2m}, \ldots, \sigma_m^2)'$.

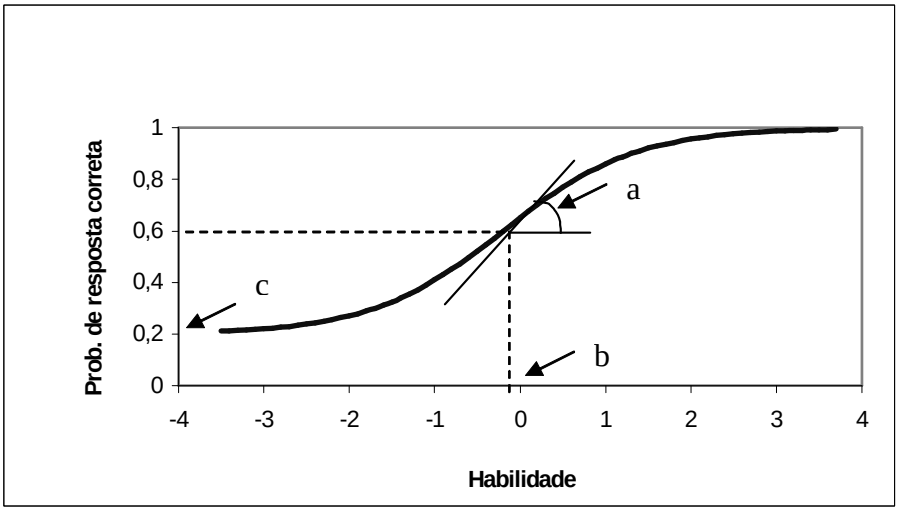
2.3.6 Estrutura de covariância diagonal

Esta é a estrutura de covariância mais simples que existe. Ela representa o caso onde as habilidades de um dado respondente ao longo do tempo são independentes, um padrão que não se espera no nosso estudo, neste caso, as covariâncias são todas iguais a zero. Essa estrutura de covariância é representada por

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_m^2 \end{pmatrix},$$

onde $\Psi = (\sigma_1^2, \ldots, \sigma_m^2)'$.

Neste trabalho serão consideradas apenas dois tipos de matrizes de covariância: uniforme e não-estruturada.


Prob. de resposta correta
1
0,8
0,6
0,4
0,2
0
a
c
b
-4
-3
-2
-1
0
1
2
3
4
Habilidade

Capítulo 3

Estimação Bayesiana

3.1 Introdução

Na estatística, toda informação que se tem sobre uma certa quantidade de interesse ϕ (não observável), se torna fundamental para continuação do estudo. Devido o verdadeiro valor de ϕ ser desconhecido, o objetivo é tentar reduzir esta nossa falta de conhecimento. Além disso, a intensidade do desconhecimento a respeito de ϕ pode assumir diferentes graus. Na inferência bayesiana, estes diferentes graus de desconhecimento são representados através da adoção de modelos probabilísticos para ϕ. Assim, é natural que o grau de conhecimento a cerca de ϕ não seja o mesmo para todos os pesquisadores, ou seja, é comum a especificação de modelos distintos.

Considere que a informação de que dispomos sobre uma certa quantidade de interesse desconhecida ϕ, resumida probabilisticamente através de $\pi(\phi)$, pode ser aumentada observando-se uma quantidade aleatória Θ relacionada com ϕ. A distribuição amostral $p(\theta|\phi)$ define esta relação. A idéia de que após observarmos $\Theta = \theta$, a quantidade de informação a respeito de ϕ aumenta, é bastante intuitiva e o *teorema de Bayes* é a regra de atualização utilizada para quantificar este aumento de informação,

$$p(\phi|\theta) = \frac{p(\theta|\phi)\pi(\phi)}{p(\theta)} = \frac{p(\theta|\phi)\pi(\phi)}{\int p(\phi,\theta)d\phi} \tag{3.1}$$

onde $\left[\int p(\phi,\theta)d\phi\right]^{-1}$ é uma constante.

Para um valor fixo de θ, a função $L(\phi) = p(\theta|\phi)$ fornece a *plausibilidade ou verossimilhança* de cada um dos possíveis valores de ϕ, enquanto $\pi(\phi)$ é chamada *distribuição a priori* de ϕ. Estas duas fontes de informação, priori e verossimilhança, são combinadas levando à *distribuição posteriori* de ϕ, $p(\phi|\theta)$, a qual será usada para fazer inferências. Assim, a forma usual do teorema de Bayes é

$$p(\phi|\theta) \propto L(\phi) \times \pi(\phi) \tag{3.2}$$

A estimação bayesiana consiste em estimar os parâmetros de interesse com base em alguma característica da posteriori. Neste trabalho esta estimação será feita através da moda da posteriori.

3.2 Prioris

Um importante problema na análise bayesiana é como definir a distribuição a priori. Se alguma informação a priori sobre o parâmetro é avaliada, ela deverá ser incorporada a densidade a priori. Se nós não temos informação a priori, precisaremos de uma priori com influência mínima na inferência. Chamaremos esta priori de *priori não informativa*. Uma das classes mais importantes de prioris não informativas, é a das prioris de referência, primeiramente descrita por Bernardo (1979) e mais tarde desenvolvida por Berger e Bernardo (1989). O método para derivar a priori de referência é também chamado de método Berger-Bernardo. Outra classe muito importante é a das prioris de Jeffreys. Discussões mais detalhadas são encontradas em Box e Tiao (1973), Berger (1985) e Bernardo e Smith (1994).

No nosso estudo, utilizaremos uma priori de referência caso a matriz de covariância seja uniforme, e uma priori não informativa de Jeffreys caso a matriz de covariância seja não-estruturada.

3.2.1 Priori de Referência

Por definição, uma priori de referência é a priori que maximiza a informação funcional perdida do experimento. Bernardo e Smith (1994) afirmam que a priori de referência para o caso da presença de um parâmetro de pertubação é baseada no uso de uma parametrização ordenada cujos primeiro e segundo componentes são (η, λ), chamados, respectivamente, de *parâmetro de interesse* e o *parâmetro de pertubação*. A priori de referência para a parametrização ordenada (η, λ) será então construída por condicionamento resultando na forma $\pi(\eta)\pi(\lambda|\eta)$.

Quando o vetor $\boldsymbol{\phi}$ de parâmetros do modelo tem mais que dois componentes, essa idéia de sucessivos condicionamentos pode obviamente ser estendida considerando $\boldsymbol{\phi}$ como sendo parametrizado ordenadamente, digamos $(\phi_1, \ldots, \phi_k)$ e gerando, por sucessivos condicionamentos, uma priori de referência, relativa a essa parametrização ordenada, da forma

$$\pi(\boldsymbol{\phi}) = \pi(\phi_1)\pi(\phi_2|\phi_1)\ldots\pi(\phi_k|\phi_1, \ldots, \phi_{k-1}). \tag{3.3}$$

Para descrever o algoritmo para produzir essa forma condicionada sucessivamente, primeiramente precisamos introduzir alguma notação.

Assumindo o modelo paramétrico $p(\theta|\boldsymbol{\phi})$, $\boldsymbol{\phi} \in \boldsymbol{\Phi}$, de modo que a matriz de informação de Fisher

$$H(\boldsymbol{\phi}) = -E_{\theta|\boldsymbol{\phi}}\left\{\frac{\partial^2}{\partial\phi_i\partial\phi_j}\ln p(\theta|\boldsymbol{\phi})\right\} \tag{3.4}$$

tenha posto completo, definimos:

1. $S(\boldsymbol{\phi}) = H^{-1}(\boldsymbol{\phi})$

2. os vetores componentes $\boldsymbol{\phi}^{[j]} = (\phi_1, ..., \phi_j)$, $\boldsymbol{\phi}_{[j]} = (\phi_{j+1}, \ldots, \phi_k)$,

e denotemos por $S_j(\boldsymbol{\phi})$ a correspondente sub-matriz superior esquerda $j \times j$ de $S(\boldsymbol{\phi})$, e por $h_j(\boldsymbol{\phi})$ o elemento inferior direito de $S_j^{-1}(\boldsymbol{\phi})$.

Finalmente, assumiremos que $\boldsymbol{\Phi} = \Phi_1 \times \cdots \times \Phi_k$, com $\phi_i \in \Phi_i$, i=1,2,...,k, denotemos por $\left\{\Phi_i^l\right\}$, l=1,2,..., uma sequência crescente de subconjuntos compactos de Φ_i, e definamos $\boldsymbol{\Phi}_{[j]}^l = \Phi_{j+1}^l \times \cdots \times \Phi_k^l$.

Proposição 1. *Com a notação acima, e sob as condições de regularidade, a priori de referência $\pi(\boldsymbol{\phi})$, relativa a parametrização ordenada $(\phi_1, \ldots, \phi_k)$, é dada por*

$$\pi(\boldsymbol{\phi}) = \lim_{l \to \infty} \frac{\pi^l(\boldsymbol{\phi})}{\pi^l(\boldsymbol{\phi}^*)}, \quad \textit{para} \quad \textit{algum} \quad \boldsymbol{\phi}^* \in \boldsymbol{\Phi}, \tag{3.5}$$

onde $\pi^l(\boldsymbol{\phi})$ é definida pelas seguintes recursões:

1. *Para $j = k$, e $\phi_k \in \Phi_k^l$,*

$$\pi_k^l\left(\boldsymbol{\phi}_{[k-1]} | \boldsymbol{\phi}^{[k-1]}\right) = \pi_k^l(\phi_k | \phi_1, \cdots, \phi_{k-1}) = \frac{\left\{h_k(\boldsymbol{\phi})\right\}^{1/2}}{\int_{\Phi_k^l} \left\{h_k(\boldsymbol{\phi})\right\}^{1/2} d\phi_k}. \tag{3.6}$$

2. *Para $j = k-1, k-2, ..., 2$, e $\phi_j \in \Phi_j^l$,*

$$\pi_j^l\left(\boldsymbol{\phi}_{[j-1]} | \boldsymbol{\phi}^{[j-1]}\right) = \pi_{j+1}^l(\boldsymbol{\phi}_{[j]} | \boldsymbol{\phi}^{[j]}) \frac{\exp\left\{E\left[\ln\left\{h_j(\boldsymbol{\phi})\right\}^{1/2}\right]\right\}}{\int_{\Phi_j^l} \exp\left\{E\left[\ln\left\{h_j(\boldsymbol{\phi})\right\}^{1/2}\right]\right\} d\phi_j}, \tag{3.7}$$

   onde $E\left[\ln\left\{h_j(\boldsymbol{\phi})\right\}^{1/2}\right] = \int_{\Phi_{[j]}^l} \ln\left\{h_j(\boldsymbol{\phi})\right\}^{1/2} \pi_{j+1}^l\left(\boldsymbol{\phi}_{[j]} | \boldsymbol{\phi}^{[j]}\right) d\boldsymbol{\phi}_{[j]}$.

3. *Para $j = 1$, $\boldsymbol{\phi}_{[0]} = \boldsymbol{\phi}$, com $\boldsymbol{\phi}^{[0]}$ vazio, e $\pi^l(\boldsymbol{\phi}) = \pi_1^l\left(\boldsymbol{\phi}_{[0]} | \boldsymbol{\phi}^{[0]}\right)$.*

A determinação da priori de referência ordenada será simplificada se os termos $\{h_j(\boldsymbol{\phi})\}$ acima, dependerem somente de $\boldsymbol{\phi}^{[j]}$, e até mesmo grande simplificação obteremos se $H(\boldsymbol{\phi})$ for diagonal, particularmente, se, para $j = 1, ..., k$, o j-ésimo termo puder ser fatorado em um produto de uma função de ϕ_j e uma função não dependendo de ϕ_j.

Corolário 1. *Se $h_j(\boldsymbol{\phi})$ depende somente de $\boldsymbol{\phi}^{[j]}$, $j = 1, ..., k$, então*

$$\pi^l(\boldsymbol{\phi}) = \prod_{j=1}^{k} \frac{\left\{h_j(\boldsymbol{\phi})\right\}^{1/2}}{\int_{\Phi_j^l} \left\{h_j(\boldsymbol{\phi})\right\}^{1/2} d\phi_j}, \boldsymbol{\phi} \in \boldsymbol{\Phi}^l. \tag{3.8}$$

Se $H(\boldsymbol{\phi})$ é diagonal(i.e., $\phi_1, \ldots, \phi_k$, são mutuamente ortogonais), com

$$H(\boldsymbol{\phi}) = \begin{pmatrix} h_{11}(\boldsymbol{\phi}) & 0 & \cdots & 0 \\ 0 & h_{22}(\boldsymbol{\phi}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{kk}(\boldsymbol{\phi}) \end{pmatrix}$$

então $h_j(\boldsymbol{\phi}) = h_{jj}(\boldsymbol{\phi})$, $j = 1, ..., k$. Além disso, se, neste último caso,

$$\{h_{jj}(\boldsymbol{\phi})\}^{1/2} = f_j(\phi_j)g_j(\boldsymbol{\phi}),$$

onde $g_j(\boldsymbol{\phi})$ *não depende de* ϕ_j*, e se os* Φ_j^l*'s não dependem de* $\boldsymbol{\phi}$*, então*

$$\pi(\boldsymbol{\phi}) \propto \prod_{j=1}^{k} f_j(\phi_j).$$

A questão que surge obviamente é qual ordenação apropriada deve ser adotada em algum problema específico. Não existe teoria formal para guiar tal escolha, mas experiência com uma grande amplitude de exemplos sugestiona que o melhor procedimento é ordenar os componentes de $\boldsymbol{\phi}$ com base no seu interesse inferencial.

Como dito anteriormente, quando a estrutura de covariância for uniforme, adotaremos a priori de referência, então seja $\{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n\}$ uma amostra aleatória de uma distribuição normal m-variada $N_m(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$, onde $\boldsymbol{\mu}$ é da forma

1. Modelo Linear:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 = \alpha_0 + \alpha_1 t_1 \\ \vdots \\ \mu_i = \alpha_0 + \alpha_1 t_i \\ \vdots \\ \mu_m = \alpha_0 + \alpha_1 t_m \end{pmatrix} = T\boldsymbol{\alpha},$$

$$\text{onde } T = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_i \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix} \text{ e } \boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix},$$

2. Modelo Quadrático:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 = \alpha_0 + \alpha_1 t_1 + \alpha_2 t_1^2 \\ \vdots \\ \mu_i = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 \\ \vdots \\ \mu_m = \alpha_0 + \alpha_1 t_m + \alpha_2 t_m^2 \end{pmatrix} = T\boldsymbol{\alpha},$$

$$\text{onde } T = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_i & t_i^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{pmatrix} \text{ e } \boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix},$$

e

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & 1 \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}, \boldsymbol{\Psi} = \begin{pmatrix} \sigma \\ \rho \end{pmatrix}, \text{ daí } \boldsymbol{\phi} = \{\boldsymbol{\alpha}', \boldsymbol{\Psi}'\},$$

então, a sua f.d.p. será dada por

$$g(\boldsymbol{\theta}|\boldsymbol{\phi}) = (2\pi)^{-m/2} \left|\boldsymbol{\Sigma}^{-1}\right|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - T\boldsymbol{\alpha})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - T\boldsymbol{\alpha})\right\},$$

onde após aplicarmos o logaritmo, teremos

$$l(\boldsymbol{\phi}) = \ln g(\boldsymbol{\theta}|\boldsymbol{\phi}) = -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{\theta} - T\boldsymbol{\alpha})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - T\boldsymbol{\alpha}). \tag{3.9}$$

Adotando-se a seguinte parametização ordenada:

- $\boldsymbol{\phi} = \{\rho, \sigma^2, \alpha_1, \alpha_0\}$, para o modelo linear
- $\boldsymbol{\phi} = \{\rho, \sigma^2, \alpha_2, \alpha_1, \alpha_0\}$, para o modelo quadrático

após alguns cálculos (ver Apêndice) obtém-se a matriz de informação de Fisher H com a seguinte forma

$$H = \begin{bmatrix} B & \mathbf{0} \\ \mathbf{0} & C \end{bmatrix} \tag{3.10}$$

cuja inversa será dada por

$$H^{-1} = \begin{bmatrix} B^{-1} & \mathbf{0} \\ \mathbf{0} & C^{-1} \end{bmatrix}$$

onde

$$B = \begin{bmatrix} \dfrac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2} & \dfrac{-m(m-1)\rho}{2\sigma^2(1-\rho)[1+(m-1)\rho]} \\ \dfrac{-m(m-1)\rho}{2\sigma^2(1-\rho)[1+(m-1)\rho]} & \dfrac{m}{2\sigma^4} \end{bmatrix}$$

$$\Rightarrow |B| = \frac{m^2(m-1)}{4\sigma^4(1-\rho)^2[1+(m-1)\rho]^2}$$

$$\Rightarrow B^{-1} = \begin{bmatrix} \dfrac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)} & \dfrac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} \\ \dfrac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & \dfrac{2\sigma^4[1+(m-1)\rho^2]}{2m} \end{bmatrix}$$

Para o modelo linear a matriz C é da forma

$$C = \begin{bmatrix} \dfrac{[1+(m-1)\rho]S_2 - \rho S_1^2}{\sigma^2(1-\rho)[1+(m-1)\rho]} & \dfrac{S_1}{\sigma^2[1+(m-1)\rho]} \\ \dfrac{S_1}{\sigma^2[1+(m-1)\rho]} & \dfrac{m}{\sigma^2[1+(m-1)\rho]} \end{bmatrix}$$

então, sua matriz inversa será dada por

$$C^{-1} = \begin{bmatrix} \dfrac{m\sigma^2(1-\rho)}{S} & \dfrac{-\sigma^2(1-\rho)S_1}{S} \\ \dfrac{-\sigma^2(1-\rho)S_1}{S} & \dfrac{\sigma^2\left\{[1+(m-1)\rho]S_2 - \rho S_1^2\right\}}{S} \end{bmatrix}$$

onde, S_1, S_2 e S são funções apenas de T. Portanto, aplicando-se o algorítmo, e sendo $S(\boldsymbol{\phi}) = H^{-1}$, obtém-se

$$S_1(\boldsymbol{\phi}) = \frac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)}$$

$$\Rightarrow S_1^{-1}(\boldsymbol{\phi}) = \frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2}$$
$$\Rightarrow h_1(\boldsymbol{\phi}) = \frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2}$$

$S_2(\boldsymbol{\phi}) = B^{-1}$

$$\Rightarrow S_2^{-1}(\boldsymbol{\phi}) = B$$
$$\Rightarrow h_2(\boldsymbol{\phi}) = \frac{m}{2\sigma^4}$$

$$S_3(\boldsymbol{\phi}) = \begin{bmatrix} \frac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)} & \frac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & 0 \\ \frac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & \frac{2\sigma^4[1+(m-1)\rho^2]}{2m} & 0 \\ 0 & 0 & \frac{m\sigma^2(1-\rho)}{S} \end{bmatrix}$$

$$\Rightarrow S_3^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2} & \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} & 0 \\ \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} & \frac{m}{2\sigma^4} & 0 \\ 0 & 0 & \frac{S}{m\sigma^2(1-\rho)} \end{bmatrix}$$

$$\Rightarrow h_3(\boldsymbol{\phi}) = \frac{S}{m\sigma^2(1-\rho)}$$

$S_4(\boldsymbol{\phi}) = H^{-1}$

$$\Rightarrow S_4^{-1}(\boldsymbol{\phi}) = H$$
$$\Rightarrow h_4(\boldsymbol{\phi}) = \frac{m}{\sigma^2[1+(m-1)\rho]}$$

Como $h_j(\boldsymbol{\phi})$ depende apenas de $\boldsymbol{\Phi}^{[j]}, j = 1, \ldots, 4$, então

$$\pi^l(\boldsymbol{\phi}) = \prod_{j=1}^{4} \frac{\{h_j(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_j^l} \{h_j(\boldsymbol{\phi})\}^{1/2} d\phi_j}, \boldsymbol{\phi} \in \boldsymbol{\Phi}^l, l \in \{1, 2, 3, \cdots\}. \tag{3.11}$$

Portanto, para a determinação da priori, devemos encontar primeiro cada um dos elementos do produto de (3.11).

$$\frac{\{h_4(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_4^l} \{h_4(\boldsymbol{\phi})\}^{1/2} d\alpha_0} = \frac{\left\{\frac{m}{\sigma^2[1+(m-1)\rho]}\right\}^{1/2}}{\int_{-e^l}^{e^l} \left\{\frac{m}{\sigma^2[1+(m-1)\rho]}\right\}^{1/2} d\alpha_0} = \frac{1}{\int_{-e^l}^{e^l} d\alpha_0} = \frac{1}{2e^l}$$

$$\frac{\{h_3(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_3^l} \{h_3(\boldsymbol{\phi})\}^{1/2} d\alpha_1} = \frac{\left\{\frac{S}{m\sigma^2(1-\rho)}\right\}^{1/2}}{\int_{-e^l}^{e^l} \left\{\frac{S}{m\sigma^2(1-\rho)}\right\}^{1/2} d\alpha_1} = \frac{1}{\int_{-e^l}^{e^l} d\alpha_1} = \frac{1}{2e^l}$$

$$\frac{\{h_2(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_2^l} \{h_2(\boldsymbol{\phi})\}^{1/2} d\sigma^2} = \frac{\left\{\frac{m}{2\sigma^4}\right\}^{1/2}}{\int_{e^{-l}}^{e^l} \left\{\frac{m}{\sigma^4}\right\}^{1/2} d\sigma^2} = \frac{\frac{1}{\sigma^2}}{\int_{e^{-l}}^{e^l} \frac{1}{\sigma^2} d\sigma^2} = \frac{1}{2l\sigma^2}$$

$$\frac{\{h_1(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_1^l} \{h_1(\boldsymbol{\phi})\}^{1/2} d\rho} = \frac{\left\{\frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2}\right\}^{1/2}}{\int_{-1+e^{-l}}^{1-e^l} \left\{\frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2}\right\}^{1/2} d\rho}$$

$$
\begin{aligned}
&= \frac{\dfrac{1}{(1-\rho)[1+(m-1)\rho]}}{\displaystyle\int_{-1+e^{-l}}^{1-e^{l}} \frac{1}{(1-\rho)[1+(m-1)\rho]} d\rho} \\
&= \frac{\dfrac{1}{(1-\rho)[1+(m-1)\rho]}}{\displaystyle\int_{-1+e^{-l}}^{1-e^{l}} \left\{ \frac{1}{m(1-\rho)} + \frac{m-1}{m[1+(m-1)\rho]} \right\} d\rho} \\
&= \frac{m}{(1-\rho)[1+(m-1)\rho] \ln \left[\dfrac{2m+(2-3m)e^{-l}+(m-1)e^{-2l}}{(2-m)e^{-l}+(m-1)e^{-2l}} \right]}
\end{aligned}
$$

Substituindo cada um dos elementos encontrados na função (3.11), a priori de referência para o modelo linear resultará em

$$\pi(\boldsymbol{\phi}) = \pi(\rho, \sigma^2, \alpha_1, \alpha_0) \propto \frac{1}{\sigma^2(1-\rho)[1+(m-1)\rho]}. \tag{3.12}$$

Para o modelo quadrático a matriz C é da forma

$$
C = \begin{bmatrix}
\dfrac{[1+(m-1)\rho]S_4 - \rho S_2^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \dfrac{[1+(m-2)\rho]S_3 - \rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \dfrac{S_2}{\sigma^2[1+(m-1)\rho]} \\
\dfrac{[1+(m-2)\rho]S_3 - \rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \dfrac{[1+(m-1)\rho]S_2 - \rho S_1^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \dfrac{S_1}{\sigma^2[1+(m-1)\rho]} \\
\dfrac{S_2}{\sigma^2[1+(m-1)\rho]} & \dfrac{S_1}{\sigma^2[1+(m-1)\rho]} & \dfrac{m}{\sigma^2[1+(m-1)\rho]}
\end{bmatrix}
$$

então, sua matriz inversa será dada por

$$
C^{-1} = \sigma^2 \begin{bmatrix}
(1-\rho)g_1(\boldsymbol{t}) & (1-\rho)g_2(\boldsymbol{t}) & (1-\rho)g_3(\boldsymbol{t}) \\
(1-\rho)g_2(\boldsymbol{t}) & (1-\rho)g_4(\boldsymbol{t}) & (1-\rho)g_5(\boldsymbol{t}) \\
(1-\rho)g_3(\boldsymbol{t}) & (1-\rho)g_5(\boldsymbol{t}) & \dfrac{g(\rho,\boldsymbol{t})}{1+(m-1)\rho}
\end{bmatrix}.
$$

onde, S_i, i=1,...,4, e $g_i(\boldsymbol{t})$, i=1,...,5, são funções apenas de T e $g(\rho, \boldsymbol{t})$ é função de ρ e T. Portanto, aplicando-se o algorítmo, e sendo $S(\boldsymbol{\phi}) = H^{-1}$, obtém-se

$$
\begin{aligned}
&S_1(\boldsymbol{\phi}) = \frac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)} \\
&\quad \Rightarrow S_1^{-1}(\boldsymbol{\phi}) = \frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2} \\
&\quad \Rightarrow h_1(\boldsymbol{\phi}) = \frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2} \\
&S_2(\boldsymbol{\phi}) = B^{-1} \\
&\quad \Rightarrow S_2^{-1}(\boldsymbol{\phi}) = B \\
&\quad \Rightarrow h_2(\boldsymbol{\phi}) = \frac{m}{2\sigma^4} \\
&S_3(\boldsymbol{\phi}) = \begin{bmatrix}
\dfrac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)} & \dfrac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & 0 \\
\dfrac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & \dfrac{2\sigma^4[1+(m-1)\rho^2]}{2m} & 0 \\
0 & 0 & \sigma^2(1-\rho)g_1(\boldsymbol{t})
\end{bmatrix}
\end{aligned}
$$

$$\Rightarrow S_3^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2} & \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} & 0 \\ \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} & \frac{m}{2\sigma^4} & 0 \\ 0 & 0 & \frac{1}{\sigma^2(1-\rho)g_1(\boldsymbol{t})} \end{bmatrix}$$

$$\Rightarrow h_3(\boldsymbol{\phi}) = \frac{1}{\sigma^2(1-\rho)g_1(\boldsymbol{t})}$$

$$S_4(\boldsymbol{\phi}) = \begin{bmatrix} \frac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)} & \frac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & 0 & 0 \\ \frac{2\sigma^2\rho(1-\rho)[1+(m-1)\rho]}{m} & \frac{\sigma^2[1+(m-1)\rho^2]}{2m} & 0 & 0 \\ 0 & 0 & \sigma^2(1-\rho)g_1(\boldsymbol{t}) & \sigma^2(1-\rho)g_2(\boldsymbol{t}) \\ 0 & 0 & \sigma^2(1-\rho)g_2(\boldsymbol{t}) & \sigma^2(1-\rho)g_4(\boldsymbol{t}) \end{bmatrix}$$

$$\Rightarrow S_4^{-1}(\boldsymbol{\phi}) = \frac{1}{\sigma^2(1-\rho)} \begin{bmatrix} \frac{m(m-1)[1+(m-1)\rho^2]\sigma^2}{2(1-\rho)[1+(m-1)\rho]^2} & \frac{-m(m-1)\rho\sigma}{[1+(m-1)\rho]} & 0 & 0 \\ \frac{-m(m-1)\rho\sigma}{[1+(m-1)\rho]} & \frac{m}{2\sigma^4} & 0 & 0 \\ 0 & 0 & \frac{g_4(\boldsymbol{t})}{[g_1(\boldsymbol{t})g_4(\boldsymbol{t})-g_2^2(\boldsymbol{t})]} & \frac{g_2(\boldsymbol{t})}{[g_1(\boldsymbol{t})g_4(\boldsymbol{t})-g_2^2(\boldsymbol{t})]} \\ 0 & 0 & \frac{g_2(\boldsymbol{t})}{[g_1(\boldsymbol{t})g_4(\boldsymbol{t})-g_2^2(\boldsymbol{t})]} & \frac{g_1(\boldsymbol{t})}{[g_1(\boldsymbol{t})g_4(\boldsymbol{t})-g_2^2(\boldsymbol{t})]} \end{bmatrix}$$

$$\Rightarrow h_4(\boldsymbol{\phi}) = \frac{g_1(\boldsymbol{t})}{\sigma^2(1-\rho)[g_1(\boldsymbol{t})g_4(\boldsymbol{t}) - g_2^2(\boldsymbol{t})]}$$

$$S_5(\boldsymbol{\phi}) = H^{-1}$$

$$\Rightarrow S_5^{-1}(\boldsymbol{\phi}) = H$$

$$\Rightarrow h_5(\boldsymbol{\phi}) = \frac{m}{\sigma^2[1+(m-1)\rho]}$$

Como $h_j(\boldsymbol{\phi})$ depende apenas de $\boldsymbol{\Phi}^{[j]}, j = 1, \ldots, 5$, então

$$\pi^l(\boldsymbol{\phi}) = \prod_{j=1}^{5} \frac{\{h_j(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_j^l} \{h_j(\boldsymbol{\phi})\}^{1/2} d\phi_j}, \boldsymbol{\phi} \in \boldsymbol{\Phi}^l, l \in \{1, 2, 3, \cdots\}. \tag{3.13}$$

Da mesma maneira como no caso linear, devemos encontar primeiro cada um dos elementos do produto de (3.13).

$$\frac{\{h_5(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_5^l} \{h_5(\boldsymbol{\phi})\}^{1/2} d\alpha_0} = \frac{\left\{\frac{m}{\sigma^2[1+(m-1)\rho]}\right\}^{1/2}}{\int_{-e^l}^{e^l} \left\{\frac{m}{\sigma^2[1+(m-1)\rho]}\right\}^{1/2} d\alpha_0} = \frac{1}{\int_{-e^l}^{e^l} d\alpha_0} = \frac{1}{2e^l}$$

$$\frac{\{h_4(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_4^l} \{h_4(\boldsymbol{\phi})\}^{1/2} d\alpha_1} = \frac{\left\{\frac{g_1(\boldsymbol{t})}{\sigma^2(1-\rho)[g_1(\boldsymbol{t})g_4(\boldsymbol{t}) - g_2^2(\boldsymbol{t})]}\right\}^{1/2}}{\int_{-e^l}^{e^l} \left\{\frac{g_1(\boldsymbol{t})}{\sigma^2(1-\rho)[g_1(\boldsymbol{t})g_4(\boldsymbol{t}) - g_2^2(\boldsymbol{t})]}\right\}^{1/2} d\alpha_1} = \frac{1}{\int_{-e^l}^{e^l} d\alpha_1} = \frac{1}{2e^l}$$

$$\frac{\{h_3(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_3^l} \{h_3(\boldsymbol{\phi})\}^{1/2} d\alpha_2} = \frac{\left\{\frac{1}{\sigma^2(1-\rho)g_1(\boldsymbol{t})}\right\}^{1/2}}{\int_{-e^l}^{e^l} \left\{\frac{1}{\sigma^2(1-\rho)g_1(\boldsymbol{t})}\right\}^{1/2} d\alpha_2} = \frac{1}{\int_{-e^l}^{e^l} d\alpha_2} = \frac{1}{2e^l}$$

$$\frac{\{h_2(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_2^l}\{h_2(\boldsymbol{\phi})\}^{1/2}\,d\sigma^2} = \frac{\left\{\frac{m}{2\sigma^4}\right\}^{1/2}}{\int_{e^{-l}}^{e^l}\left\{\frac{m}{2\sigma^4}\right\}^{1/2}d\sigma^2} = \frac{\frac{1}{\sigma^2}}{\int_{e^{-l}}^{e^l}\frac{1}{\sigma^2}d\sigma} = \frac{1}{2l\sigma^2}$$

$$\frac{\{h_1(\boldsymbol{\phi})\}^{1/2}}{\int_{\Phi_1^l}\{h_1(\boldsymbol{\phi})\}^{1/2}\,d\rho} = \frac{\left\{\frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2}\right\}^{1/2}}{\int_{-1+e^{-l}}^{1-e^l}\left\{\frac{m(m-1)}{2(1-\rho)^2[1+(m-1)\rho]^2}\right\}^{1/2}d\rho}$$

$$= \frac{\frac{1}{(1-\rho)[1+(m-1)\rho]}}{\int_{-1+e^{-l}}^{1-e^l}\frac{1}{(1-\rho)[1+(m-1)\rho]}d\rho}$$

$$= \frac{\frac{1}{(1-\rho)[1+(m-1)\rho]}}{\int_{-1+e^{-l}}^{1-e^l}\left\{\frac{1}{m(1-\rho)}+\frac{m-1}{m[1+(m-1)\rho]}\right\}d\rho}$$

$$= \frac{m}{(1-\rho)[1+(m-1)\rho]\ln\left[\frac{2m+(2-3m)e^{-l}+(m-1)e^{-2l}}{(2-m)e^{-l}+(m-1)e^{-2l}}\right]}$$

Substituindo cada um dos elementos encontrados na função (3.13), a priori de referência para o modelo quadrático resultará em

$$\pi(\boldsymbol{\phi}) = \pi(\rho, \sigma^2, \alpha_2, \alpha_1, \alpha_0) \propto \frac{1}{\sigma^2(1-\rho)[1+(m-1)\rho]}. \tag{3.14}$$

3.2.2 Priori de Jeffreys

Segundo Box and Tiao (1973) a *regra de Jeffreys* diz que a distribuição a priori para um parâmetro simples ϕ é aproximadamente não informativa se ela é proporcional a raiz quadrada da medida de informação de Fisher $I_F(\phi)$.

Para a distribuição dos parâmetros $(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$, nós assumiremos que $\boldsymbol{\alpha}$ e $\boldsymbol{\Sigma}$ são independentes, de modo que

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\Sigma}).$$

Será suposto, daqui por diante, que a parametrização em termos de $\boldsymbol{\alpha}$ é escolhida de modo que é apropriado tomarmos $\boldsymbol{\alpha}$ como localmente uniforme, ou seja

$$\pi(\boldsymbol{\alpha}) \propto \text{constante}.$$

Para a distribuição a priori dos $m(m+1)/2$ elementos distintos de $\boldsymbol{\Sigma}$, a aplicação da regra de Jeffreys para a situação multiparamétrica conduz a priori não informativa de referência

$$\pi(\boldsymbol{\Sigma}) \propto |I_F(\boldsymbol{\Sigma})|^{1|2},$$

assim,

$$\pi(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)},$$

de onde concluímos que

$$\pi(\boldsymbol{\phi}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)} \tag{3.15}$$

onde $\boldsymbol{\phi} = (\boldsymbol{\alpha}', \boldsymbol{\Psi}')$, com $\boldsymbol{\Psi}'$ sendo formado pelos $m(m+1)/2$ elementos distintos de $\boldsymbol{\Sigma}$. Para mais detalhes ver Box and Tiao (1973).

3.3 Estimação dos parâmetros

3.3.1 Verossimilhança

Façamos agora j representar um dos s padrões de resposta, $s \leq min(N, 2^{n_c})$, em vez de somente a resposta indivídual j, $r_j > 0$ ser o número de ocorrências do padrão de resposta j, j=1,2,...,s, e $\boldsymbol{R}$=$(r_1, \ldots, r_s)'$ ser o vetor (s x 1) de frequências do padrão de resposta $\boldsymbol{U}_{j..}$. Devido a independência entre as respostas dos diferentes indivíduos, $\boldsymbol{R}$ segue uma distribuição multinomial dada por

$$L(\boldsymbol{\phi}) = P(\boldsymbol{R}|\boldsymbol{\zeta}, \boldsymbol{\phi}) = \frac{N!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} \left(P(\boldsymbol{U}_{j..}|\zeta, \boldsymbol{\phi})\right)^{r_j}, \tag{3.16}$$

neste caso, a expressão (3.16) representará a função de verossimilhança, onde a probabilidade do lado direito da igualdade é dada por (2.5).

3.3.2 Equações de estimação

Dada a função de verossimilhança $L(\boldsymbol{\phi})$ e a distribuição a priori $\pi(\boldsymbol{\phi})$, então a distribuição posteriori será dada por

$$P(\boldsymbol{\phi}|\boldsymbol{u}_{jit}, \boldsymbol{\zeta}) \propto L(\boldsymbol{\phi})\pi(\boldsymbol{\phi}), \tag{3.17}$$

ou seja,

$$P(\boldsymbol{\phi}|\boldsymbol{u}_{jit}, \boldsymbol{\zeta}) \propto \frac{N!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} \left(P(\boldsymbol{U}_{j..}|\zeta, \boldsymbol{\phi})\right)^{r_j} \pi(\boldsymbol{\phi}).$$

A estimação pela moda da posteriori, consiste em obter o máximo de (3.17). Por questões de facilidade, iremos usar o logarítmo natural, que será dado por

$$\ln P(\boldsymbol{\phi}|\boldsymbol{u}_{jit}, \boldsymbol{\zeta}) = const + \ln L(\boldsymbol{\phi}) + \ln \pi(\boldsymbol{\phi}), \tag{3.18}$$

daí, o conjunto de equações de estimação para $\boldsymbol{\phi}$ será

$$\frac{\partial \ln P(\boldsymbol{\phi}|\boldsymbol{u}_{jit}, \boldsymbol{\zeta})}{\partial \boldsymbol{\phi}} = 0.$$

Primeiro observe que

$$\frac{\partial \ln P(\boldsymbol{\phi}|\boldsymbol{u}_{jit}, \boldsymbol{\zeta})}{\partial \boldsymbol{\phi}} = \frac{\partial \ln L(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} + \frac{\partial \ln \pi(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}.$$

Segundo Andrade & Tavares (2000), temos que

$$\frac{\partial \ln L(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = \sum_{j=1}^{s} r_j \int_{\Re^m} \left(\frac{\partial}{\partial \boldsymbol{\phi}} l(\boldsymbol{\phi}) \right) g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

onde

$$g_j^*(\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{\theta}|\boldsymbol{U}_{j..}, \boldsymbol{\zeta}, \boldsymbol{\phi}) = \frac{P(\boldsymbol{U}_{j..}|\boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta}|\boldsymbol{\phi})}{P(\boldsymbol{U}_{j..}|\boldsymbol{\zeta}, \boldsymbol{\phi})}. \tag{3.19}$$

Note que a probabilidade no numerador é dada por (2.4), a probabilidade no denominador por (2.5) e $g(\boldsymbol{\theta}|\boldsymbol{\phi})$ é a f.d.p. normal m-variada.

Agora, sendo $\pi(\boldsymbol{\phi})$ a priori, para determinar as suas derivadas de 1ª ordem, vamos considerar as duas estruturas de covariância propostas separadamente.

- Covariância Uniforme:

  Para o modelo com covariância uniforme, a priori é dada por (3.12), portanto

  $$\frac{\partial \ln \pi(\boldsymbol{\phi})}{\partial \boldsymbol{\alpha}} = \frac{\partial}{\partial \boldsymbol{\alpha}} \ln \left\{ \frac{1}{\sigma^2(1-\rho)[1+(m-1)\rho]} \right\} = 0$$
  $$\frac{\partial \ln \pi(\boldsymbol{\phi})}{\partial \rho} = \frac{\partial}{\partial \rho} \ln \left\{ \frac{1}{\sigma^2(1-\rho)[1+(m-1)\rho]} \right\} = \frac{2(m-1)\rho + 2 - m}{(1-\rho)[1+(m-1)\rho]}$$
  $$\frac{\partial \ln \pi(\boldsymbol{\phi})}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \ln \left\{ \frac{1}{\sigma^2(1-\rho)[1+(m-1)\rho]} \right\} = -\frac{1}{\sigma^2}$$

- Não estruturada:

  Para o modelo com covariância não estruturada, a priori é dada por (3.15), portanto

  $$\frac{\partial \ln \pi(\boldsymbol{\phi})}{\partial \boldsymbol{\alpha}} = \frac{\partial}{\partial \boldsymbol{\alpha}} \ln \left\{ |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)} \right\} = 0$$
  $$\frac{\partial \ln \pi(\boldsymbol{\phi})}{\partial \psi_i} = \frac{\partial}{\partial \psi_i} \ln \left\{ |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)} \right\} = -\tfrac{1}{2}(m+1) \frac{\partial \ln |\boldsymbol{\Sigma}|}{\partial \psi_i} = -\tfrac{1}{2}(m+1) tr \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \psi_i} \right)$$

3.3.3 Aspectos Computacionais

Para as duas estruturas de covariância consideradas neste trabalho, as equações de estimação não apresentam solução com forma fechada e portanto foi necessário aplicar algum método iterativo para obter a moda da distribuição posteriori. Neste trabalho consideramos o método de Newton-Raphson, onde sendo $\widehat{\boldsymbol{\phi}}^{(k)}$ uma estimativa para $\boldsymbol{\phi}$ na k-ésima iteração, a estimativa para $\boldsymbol{\phi}$ na *k+1*-ésima iteração será dada por

$$\widehat{\boldsymbol{\phi}}^{(k+1)} = \widehat{\boldsymbol{\phi}}^{(k)} - \left[\boldsymbol{H}(\widehat{\boldsymbol{\phi}}^{(k)}) \right]^{-1} \boldsymbol{f}(\widehat{\boldsymbol{\phi}}^{(k)}),$$

onde $\boldsymbol{f}(\boldsymbol{\phi})$ e $\boldsymbol{H}(\boldsymbol{\phi})$, são respectivamente, as matrizes de derivadas de primeira e segunda ordem do logarítmo da posteriori que é dado por (3.18) em relação a $\boldsymbol{\phi}$.

Este processo iterativo termina quando algum critério de convergência é obtido. Sendo

$$D = \left[\boldsymbol{H}(\widehat{\boldsymbol{\phi}}^{(k)}) \right]^{-1} \boldsymbol{f}(\widehat{\boldsymbol{\phi}}^{(k)})$$

um vetor (q x 1), adotou-se neste trabalho o seguinte critério, Máx$|d_{i1}| < 0.001$, $i = 1, ..., q$, onde q é igual ao número total de parâmetros do modelo.

Tanto para a simulação como para os dados reais, foi utilizado um programa de computador desenvolvido pelo prof. Dr. Héliton Tavares (co-orientador) utilizando-se a linguagem Ox (Doornik, 2000), o qual sofreu algumas modificações feitas pelo autor para a estimação dos parâmetros envolvidos nos modelos aqui propostos. Este programa está dividido da seguinte maneira:

- Corpo do Programa: nesta parte é onde ocorre a leitura das informações, leitura dos parâmetros, criação de variáveis, montagem de índices e geração de quadraturas.
- Parte Principal do Programa: nesta parte será realizada a leitura dos dados, geração das estimativas iniciais e estimação dos parâmetros da curva de crescimento quanto e dos parâmetros da matriz de covariâncias.
- Impressão dos resultados

Capítulo 4

Aplicação

4.1 Introdução

Neste capítulo aplicamos os modelos propostos em dados obtidos por simulação e em dados reais. Na Seção 4.2 trataremos da estimação dos parâmetros das curvas de crescimento, linear e quadrática, e dos parâmetros de dispersão para dados simulados. Na Seção 4.3 aplicamos a metodologia proposta a dados reais, obtidos com o projeto *PDE - Projeto de Desenvolvimento da Escola*, do Governo Federal, explorando as duas estruturas de covariância abordadas neste trabalho, uniforme e não-estruturada.

Para o estudo da simulação, consideraremos a situação onde $N = 1000$ indivíduos são submetidos a testes em $m = 5$ instantes. Para a probabilidade de acerto a qualquer item, foi utilizado o modelo Logistico com 3 parâmetros (ML3). A distribuição normal multivariada será utilizada para o vetor de habilidades. Os valores adotados para os parâmetros da distribuição das habilidades e para os parâmetros da curva de crescimento (*pcc*), são apresentados nas Tabelas 4.1 e 4.2, para as Matrizes Uniforme e não Estruturada, respectivamente. Foram escolhidos valores distintos para as médias das habilidades, mas próximos uns dos outros de modo que todos pertencessem à mesma curva de crescimento, linear ou quadrática.

Tabela 4.1 *Valores dos parâmetro da distribuição da habilidade - Matriz Uniforme*

	Médias							pcc		
	μ_1	μ_2	μ_3	μ_4	μ_5	ρ	σ^2	α_0	α_1	α_2
Linear	0	1	2	3	4	0,25	1	-1	1	
Quadrático	0	1.4	2.4	3.0	3.2	0,25	1	-1.8	2.0	-0.2

Cada um dos 5 testes será constituído por 24 itens, com 6 itens comuns entre cada par de testes adjacentes, ou seja, o número total de itens diferentes considerados nos cinco testes foi 96. Os itens serão de múltipla escolha com cinco categorias de resposta. Na Tabela 4.3 temos os valores dos parâmetros dos itens (na escala normal, isto é $D = 1,7$) utilizados na simulação. Para ambas curvas de crescimento, linear e quadrática, os valores para o parâmetro a (discriminação) variou de 0.6 (discriminação baixa) a 1.4 (discriminação alta) e os valores para o parâmetro b (dificuldade) variou de -0.7 a 4.7. Os itens com valores altos (baixos) de b foram alocados aos

Tabela 4.2 *Valores dos parâmetro da distribuição da habilidade - Matriz não Estruturada*

	Médias										pcc		
	μ_1	μ_2	μ_3	μ_4	μ_5			σ_{ij}			α_0	α_1	α_2
						1	0.5	0.4	0.3	0.2			
Linear	0	1	2	3	4		0.8	0.6	0.5	0.4	-1	1	
Quadratico	0	1.4	2.4	3.0	3.2			0.9	0.7	0.6	-1.8	2.0	-0.2
									1.1	0.8			
										1.2			

grupos onde os respondentes tem habilidades médias altas (baixas). Isto foi feito para evitar problemas na estimação. Para o parâmetro c (acerto casual) foi considerado somente um valor (0.20).

Tabela 4.3 *Valores adotados para os parâmetros dos itens*

Item	Teste	a_i	b_i	c_i	Item	Teste	a_i	b_i	c_i	Item	Teste	a_i	b_i	c_i
1	1	0.6	-0.7	0.2	33	2	1.4	1.1	0.2	65	4	1.0	3.0	0.2
2	1	1.0	-0.7	0.2	34	2	0.6	1.3	0.2	66	4	1.4	3.0	0.2
3	1	1.4	-0.7	0.2	35	2	0.6	1.5	0.2	67	4	0.6	3.1	0.2
4	1	0.6	-0.5	0.2	36	2	0.6	1.7	0.2	68	4	1.0	3.1	0.2
5	1	1.0	-0.5	0.2	37	2,3	1.0	1.3	0.2	69	4	1.4	3.1	0.2
6	1	1.4	-0.5	0.2	38	2,3	1.0	1.5	0.2	70	4	0.6	3.3	0.2
7	1	0.6	-0.3	0.2	39	2,3	1.0	1.7	0.2	71	4	0.6	3.5	0.2
8	1	1.0	-0.3	0.2	40	2,3	1.4	1.3	0.2	72	4	0.6	3.7	0.2
9	1	1.4	-0.3	0.2	41	2,3	1.4	1.5	0.2	73	4,5	1.0	3.3	0.2
10	1	0.6	-0.1	0.2	42	2,3	1.4	1.7	0.2	74	4,5	1.0	3.5	0.2
11	1	1.0	-0.1	0.2	43	3	0.6	1.9	0.2	75	4,5	1.0	3.7	0.2
12	1	1.4	-0.1	0.2	44	3	1.0	1.9	0.2	76	4,5	1.4	3.3	0.2
13	1	0.6	0.1	0.2	45	3	1.4	1.9	0.2	77	4,5	1.4	3.5	0.2
14	1	1.0	0.1	0.2	46	3	0.6	2.0	0.2	78	4,5	1.4	3.7	0.2
15	1	1.4	0.1	0.2	47	3	1.0	2.0	0.2	79	5	0.6	3.9	0.2
16	1	0.6	0.3	0.2	48	3	1.4	2.0	0.2	80	5	1.0	3.9	0.2
17	1	0.6	0.5	0.2	49	3	0.6	2.1	0.2	81	5	1.4	3.9	0.2
18	1	0.6	0.7	0.2	50	3	1.0	2.1	0.2	82	5	0.6	4.0	0.2
19	1,2	1.0	0.3	0.2	51	3	1.4	2.1	0.2	83	5	1.0	4.0	0.2
20	1,2	1.0	0.5	0.2	52	3	0.6	2.3	0.2	84	5	1.4	4.0	0.2
21	1,2	1.0	0.7	0.2	53	3	0.6	2.5	0.2	85	5	0.6	4.1	0.2
22	1,2	1.4	0.3	0.2	54	3	0.6	2.7	0.2	86	5	1.0	4.1	0.2
23	1,2	1.4	0.5	0.2	55	3,4	1.0	2.3	0.2	87	5	1.4	4.1	0.2
24	1,2	1.4	0.7	0.2	56	3,4	1.0	2.5	0.2	88	5	0.6	4.3	0.2
25	2	0.6	0.9	0.2	57	3,4	1.0	2.7	0.2	89	5	0.6	4.5	0.2
26	2	1.0	0.9	0.2	58	3,4	1.4	2.3	0.2	90	5	0.6	4.7	0.2
27	2	1.4	0.9	0.2	59	3,4	1.4	2.5	0.2	91	5	1.0	4.3	0.2
28	2	0.6	1.0	0.2	60	3,4	1.4	2.7	0.2	92	5	1.0	4.5	0.2
29	2	1.0	1.0	0.2	61	4	0.6	2.9	0.2	93	5	1.0	4.7	0.2
30	2	1.4	1.0	0.2	62	4	1.0	2.9	0.2	94	5	1.4	4.3	0.2
31	2	0.6	1.1	0.2	63	4	1.4	2.9	0.2	95	5	1.4	4.5	0.2
32	2	1.0	1.1	0.2	64	4	0.6	3.0	0.2	96	5	1.4	4.7	0.2

4.2 Estimação dos Parâmetros das Curvas de Crescimento e Dispersão

Nesta seção são apresentados alguns resultados obtidos via dados simulados, considerando cada uma das duas estruturas de covariância adotadas, uniforme e geral. As Tabelas 4.4 e 4.5 apresentam as estimativas médias e desvios-padrões para cada um dos parâmetros de interesse, onde a simulação total consistiu em 1000 replicações.

Nas Tabelas 4.4 e 4.5, observamos que as estimativas médias dos parâmetros da curva de crescimento e parâmetros de dispersão são muito próximas dos verdadeiros valores, provando a eficácia do processo de estimação dos parâmetros da curva de crescimento e parâmetros da distribuição das habilidades. Também, os desvios-padrões das estimativas dos parâmetros da curva de crescimento são muito pequenos, principalmente para o caso linear.

Tabela 4.4 *Estimativas dos Parâmetros (erro-padrão) - Dados simulados - Matriz Uniforme*

	Parâmetros da curva de crescimento			Parâmetros de Σ	
Modelo	α_0	α_1	α_2	ρ	σ
Linear	-0.9814	0.9885		0.2572	0.9995
	(0.0386)	(0.0108)		(0.0089)	(0.0005)
Quadrático	-1.7903	2.0134	-0.2021	0.2674	0.9729
	(0.0395)	(0.0336)	(0.0061)	(0.0096)	(0.0394)

4.3 Aplicação a Dados Reais

Nesta seção utilizaremos um conjunto de dados provenientes de um estudo longitudinal do Governo Federal, denominado *Projeto de Desenvolvimento da Escola (PDE)*, cujo objetivo foi acompanhar o desempenho de uma determinado grupo de estudantes, referente à habilidade nas disciplinas de Português e Matemática, quando submetidos às melhorias propostas pelo projeto. Foram selecionados indivíduos de escolas distribuídas por todo o Brasil, e esse grupo, de cerca de 15 mil crianças em abril de 1999, realizou a 1ª avaliação antes das melhorias. A partir daí foram implementadas as melhorias, e o grupo foi dividido em dois, o primeiro foi aquele em que os alunos foram submetidos as melhorias propostas pelo projeto e um outro grupo chamado controle, sendo que os dois grupos foram submetidos a mais cinco avaliações, realizadas nos meses de novembro de 1999, 2000, 2001, 2002 e 2003.

Para tornar o conjunto de dados completo e univariado, foram selecionadas todas as crianças que compareceram a todas as avaliações da disciplina matemática, totalizando 1987 crianças. Estas avaliações foram planejadas de forma a tornar comparáveis as escalas dos parâmetros dos itens e populacionais, ou seja, haviam itens comuns entre eles.

Tabela 4.5 *Estimativas dos Parâmetros (erro-padrão) - Dados simulados - Matriz não Estruturada*

Modelo	Parâmetros da Curva de Crescimento			Covariâncias				
	α_0	α_1	α_2	σ_{ij}				
Linear	-0.9975 (0.0207)	1.0078 (0.0112)		0.9883 (0.0325)	0.4829 (0.0339)	0.4451 (0.0602)	0.3575 (0.0663)	0.2063 (0.0271)
					0.9055 (0.0478)	0.6781 (0.0498)	0.5844 (0.0513)	0.4113 (0.0334)
						0.6908 (0.0458)	0.4368 (0.0407)	0.5646 (0.0399)
							1.0106 (0.0379)	0.6883 (0.0427)
								1.1401 (0.0318)
Quadrático	-1.7606 (0.0385)	1.9637 (0.0353)	-0.1958 (0.0052)	0.9879 (0.0443)	0.4921 (0.0449)	0.3961 (0.0473)	0.3129 (0.0575)	0.1898 (0.0393)
					0.9018 (0.0665)	0.6210 (0.0494)	0.5456 (0.0429)	0.3940 (0.0371)
						0.5357 (0.0537)	0.2586 (0.0680)	0.3986 (0.0758)
							0.9494 (0.1305)	0.4694 (0.0924)
								0.9883 (0.0747)

O número de itens em cada teste são, respectivamente, 34, 38, 36, 34, 40 e 34. O número de itens comuns é 10 entre os testes 1 e 2, 5 entre os testes 1 e 3, 1 entre os testes 1 e 4, 10 entre os testes 2 e 3, 4 entre os testes 2 e 4, 7 entre os testes 3 e 4, 3 entre os testes 3 e 5, 1 entre os testes 3 e 6, 10 entre os testes 4 e 5, 2 entre os testes 4 e 6, e 9 entre os testes 4 e 6 perfazendo um total de $n = 167$ itens envolvidos no estudo.

Como os parâmetros dos itens são conhecidos, será feita apenas a estimação dos parâmetros de dispersão e da curva de crescimento, considerando as duas 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 BIC (Bayesian Information Criterion) de Schwarz, que é definido como

$$BIC = lnL(\boldsymbol{\phi}; \boldsymbol{u}_{jit}, \boldsymbol{\zeta}) - \frac{c}{2} ln(n)$$

onde $\ln L(\boldsymbol{\phi}; \boldsymbol{u}_{jit}, \boldsymbol{\zeta})$ é o logarítmo natural da verossimilhança maximizada, c é o número de parâmetros, e n é o número de elementos na amostra. A estrutura de covariância que gerar o **maior** valor para BIC será a escolhida. O erro-padrão associado a cada parâmetro será estimado com base na inversa da matriz de informação de Fisher.

Nas Tabelas 4.6 e 4.7 temos as estimativas dos parâmetros populacionais, quando adotadas as curvas de crescimento linear e quadrática, respectivamente. Podemos observar que os erros-

Tabela 4.6 *Estimativas dos Parâmetros (erro-padrão) - Dados reais - Matriz Uniforme*

Modelo	Parâmetros da curva de crescimento			Parâmetros de Σ	
	α_0	α_1	α_2	ρ	σ
Linear	-0.2346	0.3478		0.5914	0.8325
	(0.0244)	(0.0031)		(0.0115)	(0.0262)
Quadrático	-0.4403	0.4934	-0.0212	0.5866	0.8220
	(0.0357)	(0.0196)	(0.0028)	(0.0117)	(0.0263)

Tabela 4.7 *Estimativas dos Parâmetros (erro-padrão) - Dados reais - Matriz não Estruturada*

Modelo	Parâmetros da CC			Parâmetros de Σ					
	α_0	α_1	α_2	σ_{ij}					
				1.0036	0.6166	0.3992	0.3293	0.6666	0.6209
				(0.0461)	(0.0353)	(0.0397)	(0.0501)	(0.0663)	(0.0543)
					0.8917	0.4112	0.3186	0.6886	0.6416
					(0.0375)	(0.0298)	(0.0503)	(0.0564)	(0.0357)
Linear	-0.4273	0.2992				0.2648	0.0652	0.4156	0.4417
	(0.0190)	(0.0047)				(0.0290)	(0.0481)	(0.0305)	(0.0298)
							0.1890	0.4055	0.4418
							(0.0491)	(0.0255)	(0.0359)
								1.0626	0.8120
								(0.0315)	(0.0287)
									1.1920
									(0.0459)
				1.0130	0.6163	0.4168	0.3129	0.6473	0.5751
				(0.0996)	(0.0636)	(0.0453)	(0.0360)	(0.0664)	(0.0604)
					0.9065	0.4333	0.3213	0.6994	0.6177
					(0.0651)	(0.0378)	(0.0284)	(0.0577)	(0.0519)
Quadrático	-0.1809	0.2101	-0,0012			0.2661	0.0625	0.4331	0.4379
	(0.0498)	(0.0219)	(0,0030			(0.0273)	(0.0162)	(0.0367)	(0.0361)
							0.1799	0.4034	0.3902
							(0.0246)	(0.0331)	(0.0292)
								1.0931	0.7386
								(0.0681)	(0.0535)
									1.0844
									(0.0608)

padrões para as estimativas dos parâmetros da curva de crescimento são pequenos, tanto considerando a covariância uniforme como a não estruturada, já em relação as estimativas dos parâmetros da matriz de covariância, temos para o caso uniforme erros-padrões bem menores que para a não estruturada.

Nas Tabelas 4.8 e 4.9 são apresentadas as log-verossimilhanças, número de parâmetros, número de elementos na amostra e os valores para BIC quando adotadas as curvas de crescimento linear e quadrática, respectivamente, que serão usados para a decisão sobre qual estrutura de covariância melhor se adapta aos dados. O maior valor obtido para BIC ocorreu quando adotada a curva de crescimento linear como matriz de covariância não estruturada. Com base nisso, optamos por

Tabela 4.8 *Log-verossimilhança associada a cada estrutura de covariância - CC Linear*

Modelo	Log-verossimilhança	c	n	BIC
Uniforme	-246.633,65	4	1987	-246.648,84
Não Estruturada	-245.766,48	23	1987	-245.853,82

Tabela 4.9 *Log-verossimilhança associada a cada estrutura de covariância - CC Quadrática*

Modelo	Log-verossimilhança	c	n	BIC
Uniforme	-248.984,94	5	1987	-249.003,93
Não Estruturada	-246.427,73	24	1987	-246.518,86

eleger o modelo linear para a curva de crescimento com matriz de covariância não estruturada, como o que melhor se ajustou aos dados.

Capítulo 5

Conclusões

Neste trabalho, assumiu-se uma distribuição Normal Multivariada para as habilidades, mas em algumas situações isto pode não ser adequado, pois a condição de simetria não é assegurada. Por exemplo, quando os indivíduos envolvidos no estudo foram pré-selecionados em uma primeira etapa, tal como ocorre em exames seriados de vestibulares. Nestes casos, devem ser propostos modelos de probabilidades alternativos. Foram exploradas apenas duas estruturas de covariância, uniforme e não estruturada, mas em muitos casos estas estruturas podem não ser adequadas.

Também foi adotado um particular modelo para explicar a função de resposta ao item, a Logística de três parâmetros, que também carrega uma característica de simetria. Funções de resposta assimétrica, tal como a *skew normal* podem ser mais adequadas em alguns casos.

Observou-se que a metodologia Bayesiana foi um poderoso instrumento no processo de estimação dos parâmetros, uma vez que produziu estimativas coerentes com o espaço paramétrico. Esta característica da Inferência Bayesiana é importante em processos de estimação onde existem parâmetros com restrições, como foi o caso dos parâmetros de dispersão aqui estudados. Nestes casos, a Inferência Clássica pode produzir estimativas que não amarrem o suporte dos parâmetros.

Em relação aos resultados da aplicação da metodologia, verificamos que os resultados foram bastante satisfatórios, onde foram produzidas estimativas bem precisas, tanto para os dados simulados como para os dados reais. Em relação aos dados do PDE, temos a ressaltar principalmente as diferenças entre as variâncias das habilidades dos alunos em cada instante, onde podemos observar que não existe uma diferença acentuada entre as variâncias nos testes 1, 2, 5 e 6, o mesmo ocorrendo com as variâncias nos testes 3 e 4. Agora, se considerarmos a diferença entre os dois grupos, teremos uma diferença acentuada, o que reforça a adoção da matriz de covariância não estruturada.

Em pesquisas futuras, pretende-se implementar este processo de estimação em outros modelos da TRI, e considerar outras estruturas de Covariância, por exemplo, a AR(1) e a Heterocedástica. Outros modelos também podem ser considerados para as curvas de crescimento, e combinações dessas propostas também podem estar associadas ao caso em que temos várias populações em estudo (Múltiplos Grupos), bem como ao caso em que temos mais de uma área do conhecimento sendo medida (Multivariada) e/ou o caso em que temos mais de uma proficiência influindo

na probabilidade de acerto ao item (Multidimensional). Pode-se também considerar a implementação da estimação conjunta dos parâmetros da distribuição das habilidades e parâmetros dos itens.

Apêndice A

Desenvolvimento das Expressões do Capítulo 3

A.1 Expressões básicas

Primeiramente vamos fazer algumas notações:

$$S_p = \sum_{i=1}^{m} t_i^p$$

$$S_{pp} = \sum_{1\leq i<j\leq m} t_i^p t_j^p$$

$$S_{ppp} = \left(\sum_{1\leq i<j<k\leq m} t_i^p t_j^p t_k^p \right) I_{\{3,4,\ldots\}}(m)$$

$$S_{pq} = \sum_{\substack{i,j=1 \\ i\neq j}}^{m} t_i^p t_j^q$$

$$S_{pqr} = \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{m} t_i^p t_j^q t_k^r$$

com $1 \leq p, q, r \leq 9$.

As expressões a seguir foram utilizadas nas demonstrações do capítulo 3:

$$S_p^2 = S_{2p} + 2S_{pp}$$

$$S_1 S_2 = S_3 + S_{21}$$

$$S_2 S_4 = S_6 + S_{42}$$

$$S_1^2 S_4 = S_6 + S_{42} + S_{51} + 2S_{411}$$

$$S_2^3 = S_6 + 2S_{42} + 6S_{222}$$

$$S_3 S_{21} = S_{51} + S_{42} + S_{321}$$

$$S_{21}^2 = S_{42} + 2S_{33} + 2S_{411} + 6S_{222}$$

$$S_1^2 S_2^2 = S_6 + 2S_{51} + 3S_{42} + 4S_{33} + 2S_{411} + 4S_{321} + 6S_{222}$$

$$S = mS_2 - S_1^2$$

A.2 Derivadas parciais

Para chegarmos a expressão de $H(\boldsymbol{\phi})$ dada por (3.10), precisaremos primeiramente determinar as derivadas parciais de primeira e segunda ordem de (3.9) em relação aos componentes de $\boldsymbol{\phi}$ e os respectivos simétricos de suas esperanças, o que será feito logo a seguir.

Derivando (3.9) em relação a $\boldsymbol{\alpha}$, temos

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\alpha}}l(\boldsymbol{\phi}) &= -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\alpha}}\left[(\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right]\\ &= -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\alpha}}\left[-2T^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right]\\ &= T^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-T\boldsymbol{\alpha}).\end{aligned} \tag{A.1}$$

cujas derivadas parciais de 2ª ordem são dadas por

$$\begin{aligned}\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^t}l(\boldsymbol{\phi}) &= \frac{\partial}{\partial\boldsymbol{\alpha}}T^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\\ &= -\frac{\partial}{\partial\boldsymbol{\alpha}}T^t\boldsymbol{\Sigma}^{-1}T\boldsymbol{\alpha}\\ &= -\left(T^t\boldsymbol{\Sigma}^{-1}T\right)^t\\ &= -T^t\boldsymbol{\Sigma}^{-1}T\end{aligned} \tag{A.2}$$

e

$$\begin{aligned}\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\psi_i}l(\boldsymbol{\phi}) &= \frac{\partial}{\partial\psi_i}T^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\\ &= T^t\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\psi_i}(\boldsymbol{\theta}-T\boldsymbol{\alpha}).\end{aligned} \tag{A.3}$$

Derivando (3.9) em relação a cada um dos componentes de $\boldsymbol{\Psi}$, encontramos

$$\begin{aligned}\frac{\partial}{\partial\psi_i}l(\boldsymbol{\phi}) &= -\frac{1}{2}\frac{\partial}{\partial\psi_i}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\frac{\partial\Sigma^{-1}}{\partial\psi_i}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\\ &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right) + (\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\psi_i}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right]\end{aligned} \tag{A.4}$$

cujas derivadas parcias de 2ª ordem são dadas por

$$\begin{aligned}\frac{\partial^2}{\partial\psi_i\partial\boldsymbol{\alpha}}l(\boldsymbol{\phi}) &= \frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\psi_i}l(\boldsymbol{\phi})\\ &= T^t\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\psi_i}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\end{aligned} \tag{A.5}$$

e

$$\begin{aligned}\frac{\partial^2}{\partial\psi_i\partial\psi_j}l(\boldsymbol{\phi}) &= \frac{\partial}{\partial\psi_j}\left\{-\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right) + (\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\psi_i}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right]\right\}\\ &= -\frac{1}{2}\left\{\frac{\partial}{\partial\psi_j}\left[tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right)\right] + (\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\psi_i\partial\psi_j}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right\}\end{aligned} \tag{A.6}$$

A seguir determinaremos os simétricos das esperanças das derivadas parciais calculadas anteriormente.

$$\begin{aligned}
-E\left\{\frac{\partial^2}{\partial\psi_i\partial\psi_j}l(\boldsymbol{\phi})\right\} &= -E\left\{-\frac{1}{2}\left\{\frac{\partial}{\partial\psi_j}\left[Tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right)\right]+(\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\psi_i\partial\psi_j}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right\}\right\} \\
&= \frac{1}{2}\left\{\frac{\partial}{\partial\psi_j}\left[Tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right)\right]+E\left[(\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\psi_i\partial\psi_j}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right]\right\} \\
&= \frac{1}{2}\left\{\frac{\partial}{\partial\psi_j}\left[Tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right)\right]+Tr\left[\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\psi_i\partial\psi_j}E\left[(\boldsymbol{\theta}-T\boldsymbol{\alpha})(\boldsymbol{\theta}-T\boldsymbol{\alpha})^t\right]\right]\right\} \\
&= \frac{1}{2}\left\{\frac{\partial}{\partial\psi_j}\left[Tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\psi_i}\right)\right]+Tr\left[\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\psi_i\partial\psi_j}\boldsymbol{\Sigma}\right]\right\}. \qquad \text{(A.7)}
\end{aligned}$$

$$\begin{aligned}
-E\left\{\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\psi_i}l(\boldsymbol{\phi})\right\} &= -E\left\{T^t\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\psi_i}(\boldsymbol{\theta}-T\boldsymbol{\alpha})\right\} \\
&= -T^t\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\psi_i}E(\boldsymbol{\theta}-T\boldsymbol{\alpha}) \\
&= \mathbf{0}_q, \qquad \text{(A.8)}
\end{aligned}$$

onde $\mathbf{0}_q$ é a matriz nula de ordem q, com q=2 para a curva de crecimento linear e q=3 para a quadrática.

$$\begin{aligned}
-E\left\{\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}l(\boldsymbol{\phi})\right\} &= -E\left\{-T^t\boldsymbol{\Sigma}^{-1}T\right\} \\
&= T^t\boldsymbol{\Sigma}^{-1}T \qquad \text{(A.9)}
\end{aligned}$$

A.3 Matriz inversa de Σ

Vimos anteriormente que para a obtenção de (A.7), (A.8) e (A.9) precisamos primeiro determinar $\boldsymbol{\Sigma}^{-1}$ e as suas derivadas parciais, e também $T^t\boldsymbol{\Sigma}^{-1}T$, o que será feito logo a seguir.

Seja $\boldsymbol{\Sigma}$ dada por

$$\boldsymbol{\Sigma}=\sigma^2\begin{pmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{pmatrix}=\sigma^2[(1-\rho)I_m+\rho J_m] \qquad \text{(A.10)}$$

onde I_m é a matriz identidade de ordem m e J_m é uma matriz $(m\times m)$ de uns, então, sendo $\boldsymbol{\Sigma}$ uma matriz simétrica, sua inversa também é simétrica e pode ser representada seguinte maneira

$$\boldsymbol{\Sigma}^{-1}=pI_m+qJ_m, \text{ onde } p,q\in\mathbb{R}^*.$$

Pela propriedade das matrizes inversas temos que

$$\boldsymbol{\Sigma}\times\boldsymbol{\Sigma}^{-1}=I_m,$$

portanto,

$$\sigma^2[(1-\rho)I_m + \rho J_m] \times (pI_m + qJ_m) = I_m,$$

e realizando o produto, encontramos

$$\sigma^2(1-\rho)pI_m + [\sigma^2(1-\rho)q + \rho p + m\rho q]J_m = I_m$$

daí,

$$\sigma^2(1-\rho)p = 1, \text{ o que resulta em } p = \frac{1}{\sigma^2(1-\rho)}$$

e

$$\sigma^2(1-\rho)q + \rho p + m\rho b = 0, \text{ o que resulta em } q = \frac{-\rho}{\sigma^2(1-\rho)[1+(m-1)\rho]}.$$

Então, concluímos que a matriz inversa de (A.10) é dada por

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{a\sigma^2}\left(I_m - \frac{1-a}{b}J_m\right), \tag{A.11}$$

com $a = 1-\rho$ e $b = 1+(m-1)\rho$, ou seja,

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix} 1+(m-2)\rho & -\rho & \cdots & -\rho \\ -\rho & 1+(m-2)\rho & \cdots & -\rho \\ \vdots & \vdots & \ddots & \vdots \\ -\rho & -\rho & \cdots & 1+(m-2)\rho \end{bmatrix}.$$

Sendo $\boldsymbol{\Sigma}^{-1}$ dada por (A.11), temos que $\boldsymbol{\Psi} = \{\rho, \sigma^2\}$. A seguir determinaremos as derivadas parciais de primeira e segunda ordem de (A.11) em relação aos componentes de $\boldsymbol{\Psi}$.

Derivando (A.11) em relação a σ^2, temos

$$\begin{aligned} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\left[\frac{1}{a\sigma^2}\left(Im - \frac{1-a}{b}J_m\right)\right] \\ &= -\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1} \end{aligned} \tag{A.12}$$

cujas derivadas parciais de segunda ordem de (A.12) são dadas por

$$\begin{aligned} \frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial (\sigma^2)^2} &= \frac{\partial}{\partial \sigma^2}\left[-\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}\right] \\ &= \frac{1}{\sigma^4}\boldsymbol{\Sigma}^{-1} \end{aligned} \tag{A.13}$$

e

$$\begin{aligned} \frac{\partial^2 \Sigma^{-1}}{\partial \sigma^2 \partial \rho} &= \frac{\partial}{\partial \rho}\left[-\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}\right] \\ &= -\frac{1}{\sigma^2}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \rho} \\ &= -\frac{1}{a\sigma^4}\left[\frac{1}{a}I_m + \frac{1}{b}\left(1 - \frac{1}{a} - \frac{1}{b}\right)J_m\right] \end{aligned} \tag{A.14}$$

Derivando (A.11) em relação a ρ, temos

$$\begin{aligned}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \rho} &= \frac{\partial}{\partial \rho}\left[\frac{1}{a\sigma^2}\left(Im - \frac{1-a}{b}J_m\right)\right] \\ &= \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \rho} = \frac{\partial}{\partial \rho}\left[\frac{1}{a\sigma^2}\left(I_m - \frac{1-a}{b}J_m\right)\right] \\ &= \frac{\partial}{\partial \sigma^2}\left(-\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}\right) \end{aligned} \tag{A.15}$$

cujas derivadas parciais de segunda ordem de (**??**) são dadas por

$$\begin{aligned}\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \rho^2} &= \frac{\partial}{\partial \rho}\left[\frac{\partial}{\partial \sigma^2}\left(-\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}\right)\right] \\ &= \frac{1}{\sigma^2}\left\{\frac{2}{a^3}I_m + \frac{1}{a^2b^2}[(m-1)a - b - m + 2]J_m\right\} \end{aligned} \tag{A.16}$$

e

$$\begin{aligned}\frac{\partial^2 \Sigma^{-1}}{\partial \rho \partial \sigma^2} &= \frac{\partial^2 \Sigma^{-1}}{\partial \sigma^2 \partial \rho} \\ &= -\frac{1}{\sigma^2}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \rho} \\ &= -\frac{1}{a\sigma^4}\left[\frac{1}{a}I_m + \frac{1}{b}\left(1 - \frac{1}{a} - \frac{1}{b}\right)J_m\right] \end{aligned} \tag{A.17}$$

- $\dfrac{\partial^2 \Sigma^{-1}}{\partial \rho^2} = \dfrac{\partial}{\partial \rho}\left\{\dfrac{1}{a\sigma^2}\left[\dfrac{1}{a}I_m + \dfrac{1}{b}\left(1 - \dfrac{1}{a} - \dfrac{1}{b}\right)J_m\right]\right\}$
  $= \dfrac{1}{\sigma^2}\left\{\dfrac{2}{a^3}I_m + \dfrac{1}{a^2b^2}[(m-1)a - b - m + 2]J_m\right\}$

Agora, sendo

$\delta_2^{(1)} = \dfrac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \psi_2} = \dfrac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \sigma^2}$

$= tr\left[\boldsymbol{\Sigma}^{-1}\dfrac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2}\right] = tr\left[\boldsymbol{\Sigma}^{-1}\left(-\dfrac{1}{\sigma^2}\boldsymbol{\Sigma}\right)\right]$

$= tr\left[-\dfrac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right] = tr\left[-\dfrac{1}{\sigma^2}I\right] = -\dfrac{m}{\sigma^2}$

cujas derivadas em relação a σ^2 e ρ são, respectivamente,

$\delta_{2,2}^{(2)} = \dfrac{\partial}{\partial \sigma^2}\left\{-\dfrac{m}{\sigma^2}\right\} = \dfrac{m}{\sigma^4}$ e $\delta_{2,1}^{(2)} = \dfrac{\partial}{\partial \rho}\left\{-\dfrac{m}{\sigma^2}\right\} = 0,$

$\delta_1^{(1)} = \dfrac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \psi_1} = \dfrac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \rho}$

$= tr\left[\Sigma^{-1}\dfrac{\partial \Sigma}{\partial \rho}\right] = tr\left\{\dfrac{1}{(1-\rho)[1+(m-1)\rho]}\begin{bmatrix}(m-1)\rho & -1 & \cdots & -1 \\ -1 & (m-1)\rho & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & (m-1)\rho\end{bmatrix}\right\}$

$= \dfrac{m(m-1)\rho}{(1-\rho)[1+(m-1)\rho]}$

cuja derivada em relação a ρ é,

$\delta_{1,1}^{(2)} = \dfrac{\partial}{\partial \rho}\left\{\dfrac{m(m-1)\rho}{(1-\rho)[1+(m-1)\rho]}\right\} = \dfrac{m(m-1)[1+(m-1)\rho^2]}{(1-\rho)^2[1+(m-1)\rho]^2},$

$$tr\left[\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial(\sigma^2)^2}\boldsymbol{\Sigma}\right] = tr\left[\frac{2}{\sigma^4}\boldsymbol{\Sigma}^{-1}\Sigma\right] = tr\left[\frac{2}{\sigma^4}I\right] = \frac{2m}{\sigma^4},$$

$$tr\left[\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\sigma^2\partial\rho}\boldsymbol{\Sigma}\right] = tr\left\{\frac{-1}{\sigma^2(1-\rho)[1+(m-1)\rho]}\begin{bmatrix}(m-1)\rho & 1 & \cdots & 1\\ 1 & (m-1)\rho & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & (m-1)\rho\end{bmatrix}\right\}$$

$$= -\frac{m(m-1)\rho}{\sigma^2(1-\rho)[1+(m-1)\rho]},$$

$$tr\left[\frac{\partial^2\boldsymbol{\Sigma}^{-1}}{\partial\rho^2}\boldsymbol{\Sigma}\right] = tr\left\{\frac{2(m-1)[1+(m-1)\rho^2]}{(1-\rho)^2[1+(m-1)\rho]^2}\begin{bmatrix}1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{bmatrix}\right\}$$

$$= \frac{2m(m-1)[1+(m-1)\rho^2]}{(1-\rho)^2[1+(m-1)\rho]^2},$$

e o valor de $T^t\boldsymbol{\Sigma}^{-1}T$ para o:

- Modelo Linear:

$$T^t\boldsymbol{\Sigma}^{-1}T$$
$$= \begin{pmatrix}1 & t_1\\ \vdots & \\ 1 & t_m\end{pmatrix}^t \frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix}1+(m-2)\rho & -\rho & \cdots & -\rho\\ -\rho & 1+(m-2)\rho & \cdots & -\rho\\ \vdots & \vdots & \ddots & \vdots\\ -\rho & -\rho & \cdots & 1+(m-2)\rho\end{bmatrix}\begin{pmatrix}1 & t_1\\ \vdots & \vdots\\ 1 & t_m\end{pmatrix}$$
$$= \frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix}[1+(m-1)\rho]S_2-\rho S_1^2 & (1-\rho)S_1\\ (1-\rho)S_1 & m(1-\rho)\end{bmatrix}$$

- Modelo Quadrático:

$$T^t\boldsymbol{\Sigma}^{-1}T$$
$$= \begin{pmatrix}1 & t_1 & t_1^2\\ \vdots & \vdots & \vdots\\ 1 & t_m & t_m^2\end{pmatrix}^t \frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix}1+(m-2)\rho & -\rho & \cdots & -\rho\\ -\rho & 1+(m-2)\rho & \cdots & -\rho\\ \vdots & \vdots & \ddots & \vdots\\ -\rho & -\rho & \cdots & 1+(m-2)\rho\end{bmatrix}\begin{pmatrix}1 & t_1 & t_1^2\\ \vdots & \vdots & \vdots\\ 1 & t_m & t_m^2\end{pmatrix}$$
$$= \frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix}[1+(m-2)\rho]S_4-\rho S_2^2 & [1+(m-2)\rho]S_3-\rho S_{21} & (1-\rho)S_2\\ \{1+(m-2)\rho\}S_3-\rho S_{21} & [1+(m-2)\rho]S_2-\rho S_{11} & (1-\rho)S_1\\ (1-\rho)S_2 & (1-\rho)S_1 & m(1-\rho)\end{bmatrix}$$

daí então, temos que

$$-E\left[\frac{\partial^2}{\partial(\sigma^2)^2}l(\boldsymbol{\phi})\right] = \frac{1}{2}\left[-\frac{m}{\sigma^4}+\frac{2m}{\sigma^4}\right]$$
$$= \frac{m}{2\sigma^4}$$
$$-E\left[\frac{\partial^2}{\partial\sigma^2\partial\rho}l(\boldsymbol{\phi})\right] = \frac{1}{2}\left[0-\frac{m(m-1)\rho}{\sigma^2(1-\rho)[1+(m-1)\rho]}\right]$$
$$= \frac{-m(m-1)\rho}{2\sigma^2(1-\rho)[1+(m-1)\rho]}$$
$$-E\left[\frac{\partial^2}{\partial\rho^2}l(\boldsymbol{\phi})\right] = \frac{1}{2}\left[-\frac{m(m-1)[1+(m-1)\rho^2]}{(1-\rho)^2[1+(m-1)\rho]^2}+\frac{2m(m-1)[1+(m-1)\rho^2]}{(1-\rho)^2[1+(m-1)\rho]^2}\right]$$
$$= \frac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2}$$

e

- Modelo Linear:

$$-E\left[\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^t}l(\boldsymbol{\phi})\right] = \frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix} [1+(m-1)\rho]S_2-\rho S_1^2 & (1-\rho)S_1 \\ (1-\rho)S_1 & m(1-\rho)\end{bmatrix}$$

- Modelo Quadrático:

$$-E\left[\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^t}l(\boldsymbol{\phi})\right] =$$
$$\frac{1}{\sigma^2[1+(m-1)\rho](1-\rho)}\begin{bmatrix} [1+(m-2)\rho]S_4-\rho S_2^2 & [1+(m-2)\rho]S_3-\rho S_{21} & (1-\rho)S_2 \\ \{1+(m-2)\rho\}S_3-\rho S_{21} & [1+(m-2)\rho]S_2-\rho S_{11} & (1-\rho)S_1 \\ (1-\rho)S_2 & (1-\rho)S_1 & m(1-\rho)\end{bmatrix}$$

Sendo

$$B = \begin{bmatrix} \frac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2} & \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} \\ \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} & \frac{2m}{\sigma^2}\end{bmatrix}$$

seu determinante será dado por

$$\begin{aligned}|B| &= \frac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2}\cdot\frac{2m}{\sigma^2} - \frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]}\cdot\frac{-m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} \\ &= \frac{m^2(m-1)[1+(m-1)\rho^2]}{\sigma^2(1-\rho)^2[1+(m-1)\rho]^2} - \frac{m^2(m-1)^2\rho^2}{\sigma^2(1-\rho)^2[1+(m-1)\rho]^2} \\ &= \frac{m^2(m-1)[1+(m-1)\rho^2-(m-1)\rho^2]}{\sigma^2(1-\rho)^2[1+(m-1)\rho]^2} \\ &= \frac{m^2(m-1)}{\sigma^2(1-\rho)^2[1+(m-1)\rho]^2}\end{aligned}$$

daí teremos que a sua inversa será

$$\begin{aligned}B^{-1} &= \frac{\sigma^2(1-\rho)^2[1+(m-1)\rho]^2}{m^2(m-1)}\begin{bmatrix} \frac{2m}{\sigma^2} & \frac{m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} \\ \frac{m(m-1)\rho}{\sigma(1-\rho)[1+(m-1)\rho]} & \frac{m(m-1)[1+(m-1)\rho^2]}{2(1-\rho)^2[1+(m-1)\rho]^2}\end{bmatrix} \\ &= \begin{bmatrix} \frac{2(1-\rho)^2[1+(m-1)\rho]^2}{m(m-1)} & \frac{\sigma\rho(1-\rho)[1+(m-1)\rho]}{m} \\ \frac{\sigma\rho(1-\rho)[1+(m-1)\rho]}{m} & \frac{\sigma^2[1+(m-1)\rho^2]}{2m}\end{bmatrix}\end{aligned}$$

Agora, considerando a matriz

$$C = \begin{bmatrix} \frac{[1+(m-1)\rho]S_2-\rho S_1^2}{\sigma^2(1-\rho)[1+(m-1)\rho]} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} \\ \frac{S_1}{\sigma^2[1+(m-1)\rho]} & \frac{m}{\sigma^2[1+(m-1)\rho]}\end{bmatrix}$$

seu determinante será dado por

$$\begin{aligned}|C| &= \frac{[1+(m-1)\rho]S_2-\rho S_1^2}{\sigma^2(1-\rho)[1+(m-1)\rho]}\cdot\frac{m}{\sigma^2[1+(m-1)\rho]} - \frac{S_1}{\sigma^2[1+(m-1)\rho]}\cdot\frac{S_1}{\sigma^2[1+(m-1)\rho]} \\ &= \frac{m[1+(m-1)\rho]S_2-m\rho S_1^2}{\sigma^4(1-\rho)[1+(m-1)\rho]^2} - \frac{S_1^2}{\sigma^4[1+(m-1)\rho]^2} \\ &= \frac{m[1+(m-1)\rho]S_2-m\rho S_1^2-(1-\rho)S_1^2}{\sigma^4(1-\rho)[1+(m-1)\rho]^2} \\ &= \frac{m[1+(m-1)\rho]S_2-[1+(m-1)\rho]S_1^2}{\sigma^4(1-\rho)[1+(m-1)\rho]^2}\end{aligned}$$

$$= \frac{mS_2 - S_1^2}{\sigma^4(1-\rho)[1+(m-1)\rho]}$$
$$= \frac{S}{\sigma^4(1-\rho)[1+(m-1)\rho]}$$

portanto, sua inversa será

$$C^{-1} = \frac{\sigma^4(1-\rho)[1+(m-1)\rho]}{S}\begin{bmatrix} \frac{m}{\sigma^2[1+(m-1)\rho]} & \frac{-S_1}{\sigma^2[1+(m-1)\rho]} \\ \frac{-S_1}{\sigma^2[1+(m-1)\rho]} & \frac{[1+(m-1)\rho]S_2 - \rho S_1^2}{\sigma^2(1-\rho)[1+(m-1)\rho]} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{m\sigma^2(1-\rho)}{S} & \frac{-\sigma^2(1-\rho)S_1}{S} \\ \frac{-\sigma^2(1-\rho)S_1}{S} & \frac{\sigma^2\left\{[1+(m-1)\rho]S_2 - \rho S_1^2\right\}}{S} \end{bmatrix}$$

A.4 Expressões da página 13

Sendo

$$C = \begin{bmatrix} \frac{[1+(m-1)\rho]S_4 - \rho S_2^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{[1+(m-2)\rho]S_3 - \rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{S_2}{\sigma^2[1+(m-1)\rho]} \\ \frac{[1+(m-2)\rho]S_3 - \rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{[1+(m-1)\rho]S_2 - \rho S_1^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} \\ \frac{S_2}{\sigma^2[1+(m-1)\rho]} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} & \frac{m}{\sigma^2[1+(m-1)\rho]} \end{bmatrix}$$

seu determinante será dado por

$$|C| = \frac{m\{[1+(m-1)\rho]S_4 - \rho S_2^2\}\{[1+(m-1)\rho]S_2 - \rho S_1^2\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} +$$
$$\frac{2S_1S_2\{[1+(m-2)\rho]S_3 - \rho S_{21}\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)} - \frac{S_2^2\{[1+(m-1)\rho]S_2 - \rho S_1^2\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)} -$$
$$\frac{S_1^2\{[1+(m-1)\rho]S_4 - \rho S_2^2\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)} - \frac{m\{[1+(m-2)\rho]S_3 - \rho S_{21}\}^2}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2}$$
$$= \frac{m\{[1+(m-1)\rho]^2S_2S_4 - \rho[1+(m-1)\rho]S_1^2S_4 - \rho[1+(m-1)\rho]S_2^3 + \rho^2S_1^2S_2^2\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} +$$
$$\frac{2(S_3+S_{21})\{[1+(m-2)\rho]S_3 - \rho S_{21}\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)} - \frac{[1+(m-1)\rho]S_2^3 - \rho S_1^2S_2^2}{\sigma^6[1+(m-1)\rho]^3(1-\rho)} -$$
$$\frac{[1+(m-1)\rho]S_1^2S_4 - \rho S_1^2S_2^2}{\sigma^6[1+(m-1)\rho]^3(1-\rho)} -$$
$$\frac{m\{[1+(m-2)\rho]^2S_3^2 - 2\rho[1+(m-2)\rho]S_3S_{21} + \rho^2S_{21}^2\}}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2}$$
$$= \frac{mS_2S_4 - S_1^2S_4 - S_2^3}{\sigma^6[1+(m-1)\rho](1-\rho)^2} + \frac{[2+(m-2)\rho]\rho S_1^2S_2^2}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} +$$
$$\frac{[1+(m-2)\rho]\{2-m-[m^2-2m+2]\rho\}S_3^2}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} +$$
$$\frac{\{2+(4m-8)\rho+[2m^2-6m+6]\rho^2\}S_3S_{21} - \rho[2+(m-2)\rho]S_{21}^2}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2}$$
$$= \frac{m(S_6+S_{42}) - (S_6+S_{42}+S_{51}+2S_{411}) - (S_6+2S_{42}+6S_{222})}{\sigma^6[1+(m-1)\rho](1-\rho)^2} +$$
$$\frac{[2+(m-2)\rho]\rho(S_6+2S_{51}+3S_{42}+4S_{33}+2S_{411}+4S_{321}+6S_{222})}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} +$$
$$\frac{[1+(m-2)\rho]\{2-m-[m^2-2m+2]\rho\}(S_6+2S_{33})}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} +$$

$$\frac{\{2+(4m-8)\rho+[2m^2-6m+6]\rho^2\}(S_{51}+S_{42}+S_{321})}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2} -$$
$$\frac{\rho[2+(m-2)\rho](S_{42}+2S_{33}+2S_{411}+6S_{222})}{\sigma^6[1+(m-1)\rho]^3(1-\rho)^2}$$
$$= \frac{(m+1)S_{42}+S_{51}-2(m-2)S_{33}-2S_{411}+2S_{321}-6S_{222}}{\sigma^6(1-\rho)^2[1+(m-1)\rho]}$$

Como o numerador da expressão acima é função apenas de $\boldsymbol{t} = (t_1, \ldots, t_m)$, então o representaremos simplesmente por $g(\boldsymbol{t})$, daí

$$|C| = \frac{g(\boldsymbol{t})}{\sigma^6(1-\rho)^2[1+(m-1)\rho]}. \tag{A.18}$$

Passemos agora à determinação de C^{-1}, onde para isso precisaremos primeiro calcular a matriz cofatora de C, a qual chamaremos de K, cujos elementos são calculados a seguir.

$$k_{11} = \begin{vmatrix} \frac{[1+(m-1)\rho]S_2-\rho S_1^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} \\ \frac{S_1}{\sigma^2[1+(m-1)\rho]} & \frac{m}{\sigma^2[1+(m-1)\rho]} \end{vmatrix}$$
$$\Rightarrow k_{11} = \frac{1}{\sigma^4[1+(m-1)\rho]^2(1-\rho)^2} \begin{vmatrix} [1+(m-1)\rho]S_2-\rho S_1^2 & (1-\rho)S_1 \\ (1-\rho)S_1 & m(1-\rho) \end{vmatrix}$$
$$= \frac{S}{\sigma^4[1+(m-1)\rho](1-\rho)}$$

$$k_{12} = \begin{vmatrix} \frac{[1+(m-2)\rho]S_3-\rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} \\ \frac{S_2}{\sigma^2[1+(m-1)\rho]} & \frac{m}{\sigma^2[1+(m-1)\rho]} \end{vmatrix}$$
$$\Rightarrow k_{12} = \frac{1}{\sigma^4[1+(m-1)\rho]^2(1-\rho)^2} \begin{vmatrix} [1+(m-2)\rho]S_3-\rho S_{21} & (1-\rho)S_1 \\ (1-\rho)S_2 & m(1-\rho) \end{vmatrix}$$
$$= \frac{(m-1)S_3-S_{21}}{\sigma^4[1+(m-1)\rho](1-\rho)}$$

$$k_{13} = \begin{vmatrix} \frac{[1+(m-2)\rho]S_3-\rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{[1+(m-1)\rho]S_2-\rho S_1^2}{\sigma^2[1+(m-1)\rho](1-\rho)} \\ \frac{S_2}{\sigma^2[1+(m-1)\rho]} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} \end{vmatrix}$$
$$\Rightarrow k_{13} = \frac{1}{\sigma^4[1+(m-1)\rho]^2(1-\rho)^2} \begin{vmatrix} [1+(m-2)\rho]S_3-\rho S_{21} & [1+(m-1)\rho]S_2-\rho S_1^2 \\ (1-\rho)S_2 & (1-\rho)S_1 \end{vmatrix}$$
$$= \frac{S_{31}-2S_{22}}{\sigma^4[1+(m-1)\rho](1-\rho)}$$

$$k_{22} = \begin{vmatrix} \frac{[1+(m-1)\rho]S_4-\rho S_2^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{S_2}{\sigma^2[1+(m-1)\rho]} \\ \frac{S_2}{\sigma^2[1+(m-1)\rho]} & \frac{m}{\sigma^2[1+(m-1)\rho]} \end{vmatrix}$$
$$\Rightarrow k_{22} = \frac{1}{\sigma^4[1+(m-1)\rho]^2(1-\rho)^2} \begin{vmatrix} [1+(m-1)\rho]S_4-\rho S_2^2 & (1-\rho)S_2 \\ (1-\rho)S_2 & m(1-\rho) \end{vmatrix}$$
$$= \frac{mS_4-S_2^2}{\sigma^4[1+(m-1)\rho](1-\rho)}$$

$$k_{23} = \begin{vmatrix} \frac{[1+(m-1)\rho]S_4-\rho S_2^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{[1+(m-2)\rho]S_3-\rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} \\ \frac{S_2}{\sigma^2[1+(m-1)\rho]} & \frac{S_1}{\sigma^2[1+(m-1)\rho]} \end{vmatrix}$$
$$\Rightarrow k_{23} = \frac{1}{\sigma^4[1+(m-1)\rho]^2(1-\rho)^2} \begin{vmatrix} [1+(m-1)\rho]S_4-\rho S_2^2 & [1+(m-2)\rho]S_3-\rho S_{21} \\ (1-\rho)S_2 & (1-\rho)S_1 \end{vmatrix}$$

$$= \frac{S_{41} - S_{32}}{\sigma^4[1+(m-1)\rho](1-\rho)}$$

$$k_{33} = \begin{vmatrix} \frac{[1+(m-1)\rho]S_4 - \rho S_2^2}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{[1+(m-2)\rho]S_3 - \rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} \\ \frac{[1+(m-2)\rho]S_3 - \rho S_{21}}{\sigma^2[1+(m-1)\rho](1-\rho)} & \frac{[1+(m-1)\rho]S_2 - \rho S_1^2}{\sigma^2[1+(m-1)\rho](1-\rho)} \end{vmatrix}$$

$$\Rightarrow k_{33} = \frac{1}{\sigma^4[1+(m-1)\rho]^2(1-\rho)^2} \begin{vmatrix} [1+(m-1)\rho]S_4 - \rho S_2^2 & [1+(m-2)\rho]S_3 - \rho S_{21} \\ (1-\rho)S_2 & (1-\rho)S_1 \end{vmatrix}$$

$$= \frac{S_{41} - S_{32}}{\sigma^4[1+(m-1)\rho](1-\rho)}$$

Com os elementos calculados acima, teremos que

$$K = \frac{1}{\sigma^4[1+(m-1)\rho](1-\rho)} \begin{bmatrix} S & (m-1)S_3 - S_{21} & S_{31} - 2S_{22} \\ (m-1)S_3 - S_{21} & mS_4 - S_2^2 & S_{41} - S_{32} \\ S_{31} - 2S_{22} & S_{41} - S_{32} & S_{41} - S_{32} \end{bmatrix}$$

daí, multiplicando-se $|C|^{-1}$ por K, teremos

$$C^{-1} = \frac{\sigma^6(1-\rho)^2[1+(m-1)\rho]}{\sigma^4[1+(m-1)\rho](1-\rho)g(\boldsymbol{t})} \begin{bmatrix} S & (m-1)S_3 - S_{21} & S_{31} - 2S_{22} \\ (m-1)S_3 - S_{21} & mS_4 - S_2^2 & S_{41} - S_{32} \\ S_{31} - 2S_{22} & S_{41} - S_{32} & S_{41} - S_{32} \end{bmatrix}$$

$$= \frac{\sigma^2(1-\rho)}{g(\boldsymbol{t})} \begin{bmatrix} S & (m-1)S_3 - S_{21} & S_{31} - 2S_{22} \\ (m-1)S_3 - S_{21} & mS_4 - S_2^2 & S_{41} - S_{32} \\ S_{31} - 2S_{22} & S_{41} - S_{32} & S_{41} - S_{32} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} (1-\rho)g_1(\boldsymbol{t}) & (1-\rho)g_2(\boldsymbol{t}) & (1-\rho)g_3(\boldsymbol{t}) \\ (1-\rho)g_2(\boldsymbol{t}) & (1-\rho)g_4(\boldsymbol{t}) & (1-\rho)g_5(\boldsymbol{t}) \\ (1-\rho)g_3(\boldsymbol{t}) & (1-\rho)g_5(\boldsymbol{t}) & \frac{g(\rho,\boldsymbol{t})}{1+(m-1)\rho} \end{bmatrix}$$

onde $g_i(\boldsymbol{t}), i = 1, ..., 5$ são funções apenas de $\boldsymbol{t}$ e $g(\rho, \boldsymbol{t})$ é uma função de ρ e $\boldsymbol{t}$.

Referências Bibliográficas

[1] ANDERSON, E. B.; MADSEN, M. **Estimating the parameters of the latent population distribution**. *Psychometrika*, v. 42, p. 357-374, 1977.

[2] ANDRADE, D. F.; TAVARES, H. R., VALLE, R.C. (2000). **Teoria da Resposta ao Item**: Conceitos a Aplicações. São Paulo: Associação Brasileira de Estatística.

[3] ANDRADE, D. F.; TAVARES, H. R. (2002). **Item response theory for longitudinal data**: population parameter estimation. (**To appear in Journal of Multivariate Analysis: ISSN 0047-259X**)

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

[5] BERGER, J. O. (1985). **Statistical Decision Theory and Bayesian Analysis**. Berlim: Springer.

[6] BERNARDO, J. M. and SMITH, A. F. M. (1994). **Bayesian Theory**. New York: John Wiley & Sons.

[7] BOCK, R. D. and LIEBERMAN, M. **Fitting a response model for n dichotomously scored items**. *Psychometrika*, v. 35, p. 179-197, 1970.

[8] BOCK, R. D. and AITKIN, M. **Marginal maximum likelihood estimation of item parameters**: An application of a EM algorithm. *Psychometrika*, v. 46, p. 433-459, 1981.

[9] BOCK, R. D. and ZIMOWSKI, M. F. (1997). **Multiple Group IRT**. In W.J. van der Linder e R.K. Hambleton Eds. *Handbook of Modern Item Response Theory*. New York: Spring-Verlag.

[10] BOX, G. E. P. and TIAO, G. C. (1973) **Bayesian Inference in Statistical Analysis**. Reading, MA: Addison-Wesley.

[11] DAVIDSON, R.; MACKINNON, J. G. (1993). **Estimation and Inference in University Econometrics**. New York: Oxford University Press.

[12] DE LEEUW, J.; VERHELST, N. **Maximum likelihood estimation in generalized Rasch Models**. *Jounal of Educational Statistics*, v. 11, p. 193-196, 1986.

[13] 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.

[14] GRAYBILL, F. A. (1969). **Introduction to Matrices with Aplications in Statistics**. Belmont, CA: Wadsworth Publishing Company, Inc.

[15] HAMBLETON, R. K.; SWAMINATHAN, H. and ROGERS, H. J. (1991). **Fundamentals of Item Response Theory**. Newbury Park: Sage Publications.

[16] LORD, F. M. (1952) **A theory of test scores (No. 7).**. *Psychometrika Monograph.*

[17] LORD, F. M. and Novick, M. R. (1968) **Statistical Theories of Mental Test Scores**. Reading, MA: Addison-Wesley.

[18] MISLEVY, R. J. **Estimating latent distributions**. *Psychometrika*, v. 49, p. 359-381, 1984.

[19] RASCH, G. (1960) **Probabilistic Models for Some Intelligence and Attainment Test**. Copenhagen: Danish Institute for Educational Research.

[20] SANATHANAN, L.; BLUMENTHAL, N. **The logistic model and estimation of latent structure**. *Journal of the American Statistical Association*, v. 73, p. 794-798, 1978.

[21] SEARLE, S. R. (1982) **Matrix Algebra Useful for Statistics**. New York: John Wiley & Sons.

[22] SINGER, J. M.; ANDRADE, D. A. **Analysis of longitudinal data**. In P. K. Sen & C. R. Rao, Eds. *In Handbook of Statistics*, v. 18, p. 115-160, 2000.