



**Universidade de Brasília**  
**Instituto de Ciências Exatas**

Departamento de Estatística

Dissertação de Mestrado

**Uso de Regressão Isotônica na Escolha de Itens  
em Testes Adaptativos Computadorizados**

por

**Alfredo Moreira Salgado**

Orientador: Prof. Dr. Antônio Eduardo Gomes

Co-Orientador: Prof. Dr. Dalton Francisco de Andrade

**Julho de 2018**

# Uso de Regressão Isotônica na Escolha de Itens em Testes Adaptativos Computadorizados

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Comissão Julgadora:

- Prof. Dr. Antonio Eduardo Gomes (orientador) - EST/UnB
- Prof. Dr. Dalton Francisco de Andrade (coorientador) - DIE/UFSC
- Prof. Dr. Raul Yukihiro Matsushita (membro interno) - EST/UnB
- Prof. Dr. Caio Lucidius Naberezny Azevedo (membro externo) - IMECC/Unicamp
- Profa. Dra. Cibele Queiroz da Silva (suplente) - EST/UnB

**Universidade de Brasília**

**Brasília, Julho de 2018**

*Aos meus falecidos avôs, Fábio e Vicente, que estariam felizes com a conclusão deste mestrado.*

# Agradecimentos

Agradeço a Deus, primeiramente, pois por Ele, e para Ele, são todas as coisas. Agradeço, também, aos meus familiares que tanto me apoiaram nestes dois anos de mestrado, sem deixar de nominar meus pais, Alfredo Cláudio e Marluce; minha irmã Rebeca e meu irmão Davi; e minha amada esposa, Isabelle; além de cunhados, madastra, tios, primos, avós, sogros e todos que apoiaram minha decisão de concluir o mestrado.

Aos meus amigos, que me incentivaram insistentemente para conclusão desta dissertação e, principalmente, aqueles que oraram para que essa etapa fosse concluída.

Agradeço a meu orientador, professor Antônio Eduardo, bem como o co-orientador, professor Dalton, pela paciência, conselhos e auxílio neste trabalho. Às bancas de qualificação e defesa; compostas pelos professores Caio, Raul e Cibele; pelas ponderações e ajustes propostos para melhor redação do trabalho final.

Por fim, agradeço aos meus colegas de trabalho que entenderam minha ausência em alguns momentos, especialmente na licença capacitação tirada, e minha priorização pelo término do curso, em especial meus chefes Gleidson e Gustavo, minha ex-chefe Greiceane e meu colega Juciê.

# Sumário

<b>Lista de Figuras</b>	<b>4</b>
<b>Lista de Tabelas</b>	<b>6</b>
<b>Lista de Abreviaturas</b>	<b>7</b>
<b>Resumo</b>	<b>8</b>
<b>Abstract</b>	<b>9</b>
<b>Introdução</b>	<b>10</b>
<b>1 Referencial Teórico</b>	<b>12</b>
1.1 Curva Característica do Item . . . . .	13
1.2 Função de Informação . . . . .	14
1.3 Métrica de Traço Latente e Suposições do Modelo . . . . .	15
1.4 Calibração e Estimação . . . . .	16
1.5 Outros Modelos . . . . .	17
1.6 Testes Adaptativos Computadorizados . . . . .	17
1.6.1 Estimativa do traço latente . . . . .	18
1.6.2 Critérios de Seleção do Item . . . . .	21
1.7 Regressão Isotônica . . . . .	25
1.7.1 Algoritmo PAV . . . . .	29
1.7.2 Método de Nadaraya-Watson . . . . .	30
1.8 Considerações . . . . .	30
<b>2 Metodologia Proposta</b>	<b>32</b>
2.1 Simulações . . . . .	35

2.2 Banco de Itens Real . . . . .	50
<b>3 Conclusão e Propostas</b>	<b>53</b>
A Quadratura Gaussiana	59
B Tabelas Resumo da Simulação	61

# Lista de Figuras

1.1	Curva Característica do Item . . . . .	14
1.2	Seleção do item no nível de traço latente impreciso. . . . .	22
1.3	Representação gráfica do Diagrama de Soma Acumulada (DSA) e a função Minorante Convexa Máxima (MCM) . . . . .	27
2.1	Escolha do próximo item com uso da Regressão Antitônica suavizada por Nadaraya-Watson ( $h=1$ ) . . . . .	34
2.2	Quantidade média de itens utilizados em cada critério de escolha - Desvio Padrão a Posteriori = 0.2 . . . . .	37
2.3	Habilidade simulada x Habilidade estimada - Valores de $\theta$ positivo - Desvio Padrão a Posteriori = 0.2 . . . . .	38
2.4	Habilidade simulada x Habilidade estimada - Valores de $\theta$ negativo - Desvio Padrão a Posteriori = 0.2 . . . . .	38
2.5	Quantidade média de itens utilizados em cada critério de escolha - Desvio Padrão a Posteriori = 0.1 . . . . .	42
2.6	Habilidade simulada x Habilidade estimada - Valores de $\theta$ positivo - Desvio Padrão a Posteriori = 0.1 . . . . .	43
2.7	Habilidade simulada x Habilidade estimada - Valores de $\theta$ negativo - Desvio Padrão a Posteriori = 0.1 . . . . .	43
2.8	Histograma dos Parâmetros de Discriminação e de Dificuldade - Proficiência em Inglês da UnB . . . . .	51
2.9	Quantidade média de itens utilizados em cada critério de escolha - Profici- ência em Inglês da UnB . . . . .	51
2.10	Habilidade simulada x Habilidade estimada - Valores de $\theta$ positivo - Profi- ciência em Inglês da UnB . . . . .	52

2.11 Habilidade simulada x Habilidade estimada - Valores de $\theta$ positivo - Proficiência em Inglês da UnB . . . . .	52
---	----

# Lista de Tabelas

2.1	Tabela resumo dos critérios comparados . . . . .	36
2.2	Tabela Resumo da Simulação da Isotônica Suavizada por Nadaraya-Watson - Desvio Padrão a Posteriori = 0.2 . . . . .	39
2.3	Tabela Resumo da Simulação da Isotônica com Interpolação Linear - Desvio Padrão a Posteriori = 0.2 . . . . .	40
2.4	Tabela Resumo da Simulação da Isotônica Suavizada por Nadaraya-Watson - Desvio Padrão a Posteriori = 0.1 . . . . .	44
2.5	Tabela Resumo da Simulação da Isotônica com Interpolação Linear - Desvio Padrão a Posteriori = 0.1 . . . . .	45
2.6	Viés Médio em cada nível de habilidade - Teste de Tamanho Fixo em 30 Itens.	47
2.7	EQM em cada nível de habilidade - Teste de Tamanho Fixo em 30 Itens. . .	48
B.1	Tabela Resumo da Simulação da Máxima Informação - Desvio Padrão a Posteriori = 0.2 . . . . .	61
B.2	Tabela Resumo da Simulação da Máxima Informação Ponderada pela Pos- teriori - Desvio Padrão a Posteriori = 0.2 . . . . .	63
B.3	Tabela Resumo da Simulação da Máxima Informação Ponderada pela Ve- rossimilhança - Desvio Padrão a Posteriori = 0.2 . . . . .	65
B.4	Tabela Resumo da Simulação de Kullback-Leibler - Desvio Padrão a Poste- riori = 0.2 . . . . .	67
B.5	Tabela Resumo da Simulação da Máxima Informação - Desvio Padrão a Posteriori = 0.1 . . . . .	69
B.6	Tabela Resumo da Simulação da Máxima Informação Ponderada pela Pos- teriori - Desvio Padrão a Posteriori = 0.1 . . . . .	71
B.7	Tabela Resumo da Simulação da Máxima Informação Ponderada pela Ve- rossimilhança - Desvio Padrão a Posteriori = 0.1 . . . . .	73

B.8	Tabela Resumo da Simulação de Kullback-Leibler - Desvio Padrão a Posteriori = 0.1 . . . . .	75
-----	---	----

# Lista de Abreviaturas

ASBAV: *Armed Services Vocational Aptitude Test Battery*

CAT: *Computerized Adaptive Test*

CCI: Curva Característica do Item

Cebraspe: Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos

DSA: Diagrama de Soma Acumulada

EAP: *Expected a Posteriori*

EMV: Estimador de Máxima Verossimilhança

EQM: Erro Quadrático Médio

FIFA: *Full-Information Factorial Analysis*

GRE: *Graduate Record Examination*

MAP: *Maximum a Posteriori*

MCM: Minorante Convexo Máximo

PAV: *Poll Adjacent Violators*

TAC: Teste Adaptativo Computadorizado

TOEFL: *Test of English as a Foreign Language*

TRI: Teoria de Resposta ao Item

# Resumo

Os testes adaptativos computadorizados têm desafiado pesquisadores da área a propor diferentes métodos de estimações para o traço latente do respondente e, em particular, diversos critérios para escolha mais eficiente dos itens que serão respondidos. Neste trabalho, serão propostos dois novos critérios para seleção de item em testes adaptativos computadorizados, mediante aplicação regressão isotônica, na qual a dificuldade do item é a variável explicativa, e a indicadora de acerto é a variável resposta. Os resultados mostram que os critérios possuem resultados semelhantes aos que já são utilizados, embora se mostrem menos acurados para algumas habilidades, principalmente aquelas acima da média. A vantagem do critério que utiliza a regressão suavizada por Nadaraya-Watson é o tempo de processamento, pois chega a ser duas vezes mais rápido que o mais rápido dos outros critérios, que é o critério da Máxima Informação de Fisher. Ao final, são feitas propostas de estudos futuros na tentativa de aperfeiçoar os critérios criados, a fim de obter maiores vantagens na escolha do próximo item a ser aplicado ao respondente, reduzindo a quantidade de itens e mantendo a boa estimativa.

**Palavras-chave:** Teoria de Resposta ao Item (TRI); testes adaptativos computadorizados; regressão isotônica.

# Abstract

The Computerized Adaptive Testing have defied researchers to propose different estimation methods for the respondent's latent trait, particularly criteria for choosing more efficiently the item which will be answered. In this work we propose two new criteria for item selection in computerized adaptive tests, using isotonic regression, considering item's difficulty the explanatory variable, and the correct answer indicator, the response variable. Results are similar to already used criteria, however less efficient for some abilities is observed, especially to latent trait greater than means. The advantage of using the Nadaraya-Watson's regression smoothing is its processing time, up to twice as fast as the fastest of the other criteria, which is the maximum information. At the end, propositions for future studies are made, in an attempt to improve the created criteria, to obtain greater efficiency on choosing the next item to be assigned to the respondent, reducing the item quantity while keeping a good estimate.

**Keywords:** item response theory; computerized adaptive testing; isotonic regression.

# Introdução

A Teoria de Resposta ao Item consiste em um conjunto de ferramentas estatísticas desenvolvidas em meados da década de 50 para medir traços latentes, ou seja, características do indivíduo ou objeto (unidade observada) que não podem ser medidas de forma direta. Alguns exemplos de traços latentes são a proficiência em uma língua ou numa ciência específica.

A TRI surgiu como alternativa à Teoria Clássica da Medida na educação, principalmente por permitir a comparabilidade das habilidades estimadas ao longo de um período. Foi uma solução para as avaliações educacionais e o acompanhamento histórico da qualidade de ensino de uma determinada área. Além dessa comparabilidade ao longo do tempo, a TRI permite a comparação entre respondentes submetidos a provas com itens diferentes, já que eles também têm papel importante nos parâmetros a serem estimados.

Apesar de ter uma vasta literatura de aplicação na psicometria e na educação, a TRI já se mostrou aplicável em outras áreas como sociologia, medicina, administração de empresas, marketing e outras, como mostra Moreira Junior (2010).

Segundo Andrade et al. (2000), as vantagens da utilização da TRI dependem fundamentalmente da adequação (ajuste) dos modelos e seus pressupostos, que serão discutidos no Capítulo 1.

Com o avanço da tecnologia, criou-se a oportunidade de adaptação dos testes que já utilizavam TRI para sua implementação em computadores. Surgiram, então, os Testes Adaptativos Computadorizados (TAC) ou *Computerized Adaptive Test (CAT)*, em inglês. O CAT trouxe mais agilidade na aplicação do teste, tornando o processo menos custoso do ponto de vista humano.

Importante salientar que os testes adaptativos recebem esse nome não por serem adaptados para computadores, mas porque a escolha dos itens é adaptada ao nível do traço latente do respondente. Como explicou Wainer (2000), a ideia básica de um teste adapta-

tivo é imitar de forma automática o que um sábio examinador faria.

Os testes adaptativos, amplamente utilizados em proficiência de línguas, por exemplo, consistem em selecionar um *k*-ésimo item na aplicação de um avaliado de acordo com a proficiência estimada com as  $(k - 1)$  respostas dos outros itens; daí o termo adaptativo, pois o teste se adapta ao desempenho do usuário ao longo da aplicação.

O processo de seleção do próximo item deve ser cuidadoso, de tal forma que se aproveite, o máximo possível, as vantagens de um teste adaptativo. A aplicação de itens fáceis para um respondente com traço latente alto pode tornar tedioso o teste, permitindo que o respondente deixe de enxergar o teste como desafio e saia do estado de concentração. Por outro lado, itens muito difíceis para um respondente de baixo traço latente, o desestimula para que ele continue respondendo o teste (Sands e Waters, 1997).

Para a seleção do item subsequente, há o cuidado de evitar processos de estimação computacionalmente onerosos, o que poderia prejudicar no tempo de aplicação do teste, uma vez que a estimação é feita em tempo real à medida que os itens são respondidos. Com a capacidade de processamento atual, há menos restrição com relação à complexidade de estimação.

As vantagens enumeradas por Costa (2009) do CAT estariam relacionadas a administração e correção do teste, precisão das estimativas, segurança do teste, motivação e tecnologia. Já as desvantagens ficam por conta da maior quantidade de recursos financeiros e humanos para a sua organização.

Vários critérios de seleção do item têm sido propostos no ambiente CAT e alguns são descritos por Van der Linden e Pashley (2009). A proposta deste trabalho é propor uma nova alternativa de critério de seleção do item baseado na utilização da regressão isotônica.

O objetivo é, por meio de simulações, testar em que situações o novo critério é preferível aos critérios tradicionalmente utilizados. Para a avaliação de vários cenários, será feito um estudo experimental para distintos critérios de convergência e níveis do traço latente.

A divisão do trabalho é feita, além desta Introdução, em um capítulo trazendo todo o referencial teórico necessário para a compreensão de uma proposta apresentada no capítulo seguinte, juntamente com seus resultados de simulação e comparação com métodos já existentes, e uma conclusão dos resultados e propostas de estudos futuros.

# Capítulo 1

## Referencial Teórico

A Teoria de Resposta ao Item é uma classe de modelos que, inicialmente propostos por Lord (1952), foram aperfeiçoados e adaptados para distintas situações.

O modelo logístico de três parâmetros, proposto por Birnbaum (1968), é o mais utilizado, principalmente na área da educação. Para itens dicotômicos ou dicotomizados<sup>1</sup>, tais quais os que serão objeto de estudos neste trabalho, ele é dado por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + \exp[-Da_i(\theta_j - b_i)]}, \quad (1.1)$$

com  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, n$ , sendo:  $I$  a quantidade de itens,  $n$  o número de unidades amostrais,  $U_{ij}$  variável que assume o valor 1, caso a unidade amostral  $i$  acerte o item  $j$  e 0, caso contrário,  $\theta_j$  é o traço latente da  $j$ -ésima unidade amostral e  $D$  é o parâmetro de escala, usualmente atribuído o valor unitário ou  $D = 1,702$  quando se deseja uma aproximação da função ogiva normal.

O modelo recebe esse nome por conter três parâmetros para o item: discriminação, dificuldade e probabilidade de acerto para um respondente com baixo nível de traço latente, representados respectivamente por  $a_i$ ,  $b_i$  e  $c_i$  em (1.1). Apesar da opção de trabalhar com o modelo de três parâmetros, cabe destacar que o modelo de dois parâmetros pode ser obtido quando  $c_i = 0$  (muito utilizado em modelos psicométricos) e o modelo de um parâmetro, considerando, ainda,  $a_i = 1$  para todo  $i$ .

A equação (1.1) também é apresentada de outra forma na literatura científica, conforme

---

<sup>1</sup>Itens dicotômicos são aqueles que apenas duas respostas são possíveis, de tal forma que uma é a correta e outra é a errada. O item é chamado dicotomizado quando, apesar de existir mais de duas alternativas de resposta, é possível definir uma ou mais resposta como correta e outro grupo de respostas como incorreta

a seguir:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{\exp [Da_i(\theta_j - b_i)]}{1 + \exp [Da_i(\theta_j - b_i)]}$$

Andrade et al. (2000) demonstram que o modelo descrito em (1.1) apresenta problema de falta de identificabilidade, ou seja, diferentes valores dos parâmetros dos itens e do traço latente produzem a mesma probabilidade de acerto. Esse fato pode ser provado, na igualdade a seguir, quando um  $\theta_j^*$  é uma combinação linear de  $\theta_j$ , ou seja,  $\theta_j^* = \alpha\theta_j + \beta$ ;  $a_i^* = \frac{a_i}{\alpha}$ ;  $b_i^* = \alpha b_i + \beta$  e  $c_i^* = c_i$ .

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

Essa ausência de identificabilidade pode ser solucionada com a especificação de uma métrica de traço latente, ou seja, uma medida de posição e uma de dispersão para o traço latente .

## 1.1 Curva Característica do Item

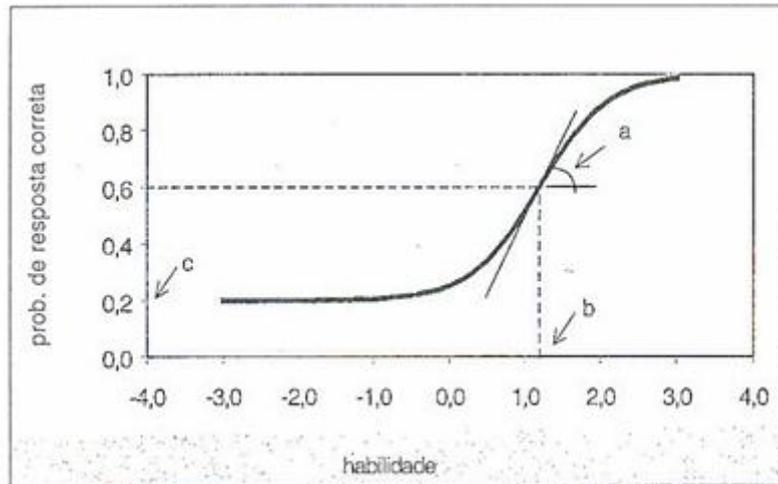
A Curva Característica do Item (CCI) é a relação gráfica entre a probabilidade de resposta correta e os parâmetros do item. A CCI é monotonicamente crescente, pois a probabilidade de acerto aumenta à medida que o traço latente aumenta e é representada na Figura 1.1.

O parâmetro de discriminação  $a$  pode assumir qualquer valor em  $\mathbb{R}$ , mas espera-se valores positivos, pois, como pode ser visto na Figura 1.1,  $a$  é proporcional à derivada da tangente no ponto de inflexão da CCI. Valores negativos de  $a$ , portanto, indicariam uma CCI decrescente em algum ponto. Valores muito baixos para  $a$  são encontrados em itens com baixa discriminação, ou seja, itens em que o traço latente do respondente não interfere tanto na probabilidade de seu acerto.

O parâmetro de dificuldade  $b$  assume valores na mesma escala da habilidade  $\theta_j$  e, por vezes, recebe o nome de habilidade do item, no ponto de inflexão,  $p_i = (c + 1)/2$ , e quanto maior for esse valor, mais difícil é o item.

Por fim, o parâmetro de probabilidade de acerto para baixo nível de traço latente,  $c$ ,

Figura 1.1: Curva Característica do Item



Fonte: Andrade, Tavares e Valle (2000)

como toda probabilidade, encontra-se entre 0 e 1. Quando  $c = 0$ , trata-se de um modelo de dois parâmetros e não tem por que se falar em acerto casual. Neste caso,  $p_i = 0.5$  no ponto de inflexão.

## 1.2 Função de Informação

Uma medida importante na TRI é a função de informação do item e a função de informação do teste. A primeira é útil na análise do quanto um item possui de informação para a estimação de um traço latente e é dada por:

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta}P_i(\theta)\right]^2}{P_i(\theta)Q_i(\theta)}, \quad (1.2)$$

sendo  $I_i(\theta)$  a informação fornecida pelo item  $i$  no traço latente  $\theta$ ,  $P_i(\theta) = P(U_{ij} = 1|\theta_j)$  e  $Q_i(\theta) = 1 - P_i(\theta)$ .

Andrade et al. (2000) descrevem a equação (1.2), para o modelo de três parâmetros, da seguinte forma:

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

Dessa forma, a equação mostra a importância dos três parâmetros na informação do item. É possível notar que a informação será maior quando  $b_i$  se aproximar de  $\theta$ , quando  $c_i$  se aproximar de 0, e quanto maior for  $a_i$ .

A função de informação do teste é outra medida muito utilizada e pode ser descrita como:

$$I(\theta) = \sum_{i=1}^I I_i(\theta). \quad (1.3)$$

A equação (1.3) fornece o erro-padrão de estimação, conforme fórmula a seguir:

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

Andrade et al. (2000) destaca, ainda, que todas essas medidas dependem do valor de  $\theta$ , ou seja, do traço latente do respondente.

### 1.3 Métrica de Traço Latente e Suposições do Modelo

Uma preocupação na TRI está relacionada à métrica do traço latente. Conforme o que está descrito na Equação (1.1), o traço latente pode assumir qualquer valor real entre  $-\infty$  e  $+\infty$ . Dessa forma, é necessário definir uma origem e uma unidade de medida para a definição da métrica. Entre os estudiosos de TRI, a métrica mais utilizada é a (0,1), ou seja, com média 0 e desvio-padrão igual a 1. Outra métrica pode ser utilizada desde que mantida as relações de ordem existentes entre os pontos e, independente da escala utilizada, a probabilidade de acerto em um determinado item deve ser invariante, pois não faz sentido a probabilidade ser alterada devido à escolha da métrica do traço latente.

São duas as suposições usualmente consideradas nos modelos de teoria de resposta ao item, em particular o modelo logístico de três parâmetros. A primeira é a unidimensionalidade do teste, que nada mais é do que a suposição de que os itens estão medindo um único traço latente ou que o traço latente avaliado seja fator dominante para a resolução dos itens propostos.

Mislevy (1986b) propôs um método baseado na máxima verossimilhança, para verificar a dimensionalidade do teste. Mais recentemente, uma análise fatorial do banco de itens usando um estimador da informação completa passou a ser utilizado para esse fim. Em inglês, o método ficou conhecido como *FIFA* (*Full-Information Factorial Analysis*) e pode ser melhor compreendido pelo trabalho de Wirth e Edwards (2007)

A segunda suposição diz respeito à independência na resposta de diferentes itens em

um mesmo nível de habilidade, também chamada de suposição da independência local. Hambleton et al. (1991) mostram que, geralmente, ao satisfazer a suposição de unidimensionalidade, a independência local já está satisfeita.

## 1.4 Calibração e Estimação

Os dois maiores desafios na TRI são a calibração dos itens, como é conhecida a etapa de estimação dos parâmetros dos itens, e estimação do traço latente. Por vezes, ambos os tipos de parâmetros (dos itens e do traço latente) são desconhecidos e outras vezes o interesse está na estimação de apenas um tipo de parâmetro, já que o outro pode ser pré-conhecido.

No geral, deseja-se estimar tanto os parâmetros dos itens quanto os traços latentes, o que torna o processo pesado, uma vez que se trata de três parâmetros para cada item e um parâmetro de traço latente para cada respondente.

De qualquer forma, quando o problema envolve a calibração dos itens e estimação do traço latente, há duas abordagens distintas: a estimação conjunta e a separada. Para as duas abordagens, há várias propostas e bibliografia que analisam as vantagens e desvantagens. Andrade et al. (2000) dedicam um capítulo inteiro para descrever algumas dessas propostas de estimação em cada abordagem aqui descrita.

Resumidamente, a maioria das propostas trata-se de usar o Método da Máxima Verossimilhança (Marginal, em alguns casos) junto a um processo iterativo como *Newton-Raphson* ou *Scoring de Fisher*. Porém, alguns problemas podem surgir principalmente em casos de poucos erros ou poucos acertos de itens e, nestes casos, recorrem-se a métodos de estimação bayesianos.

Outra questão que tem a estimação bayesiana como possível solução é quando as estimativas pelos métodos frequentistas assumem valores não esperados, como algum  $a_i$  negativo ou  $c_i$  fora do intervalo  $[0,1]$ . Esses problemas são evitados ao supor distribuições a priori para os parâmetros dos itens que condizem com o que se espera da estimativa.

A distribuição a priori de  $a_i$ , por exemplo, pode ser uma Log-Normal ou Quiquadrado. Para os  $b_i$ , adota-se uma priori Normal e para o  $c_i$  costuma-se utilizar a distribuição Beta.

Apesar da vantagem de estar definida para qualquer padrão de resposta, a estimação bayesiana tem suas desvantagens, como o fato de ser mais trabalhosa computacionalmente e a necessidade de estabelecer distribuições a priori para os parâmetros. Além disso, na estimação do traço latente, os estimadores são viesados.

## 1.5 Outros Modelos

Há, ainda, outros modelos na TRI que, dependendo do conjunto de dados, podem ou devem ser utilizados. A maioria surge de particularidades do modelo logístico de três parâmetros.

O primeiro deles é um modelo muito utilizado em testes psicométricos, onde não tem tanto valor a probabilidade de acerto para um respondente com baixo traço latente. Trata-se do modelo logístico de dois parâmetros, que pode ser obtido pela equação (1.1) para  $c_i = 0$  para todo  $i = 1, 2, \dots, I$ .

$$P_i(\theta_j) = P(U_{ij} = 1|\theta_j) = \frac{1}{1 + \exp[-Da_i(\theta_j - b_i)]}.$$

Há, ainda, o modelo de Rasch, que não leva em conta a discriminação do item, considerando todo  $a_i = 1$ . Esse modelo pode ser descrito por:

$$P_i(\theta_j) = P(U_{ij} = 1|\theta_j) = \frac{1}{1 + \exp[-(\theta_j - b_i)]}.$$

Para itens não dicotômicos ou dicotomizados, pode-se citar ainda os modelos de resposta nominal de Bock (1972), o de resposta gradual de Samejima (1969), o de escala gradual de Andrich (1978), o de crédito parcial de Masters (1982) e o de crédito parcial generalizado de Muraki (1997). Além desses, há adaptações para estudos envolvendo duas ou mais populações.

## 1.6 Testes Adaptativos Computadorizados

O avanço computacional das últimas décadas permitiu o avanço na criação de testes adaptativos no mundo inteiro. No geral, em testes adaptativos computadorizados, os parâmetros dos itens já foram estimados. Portanto, a parte computacional é exigida, principalmente, para a estimativa parcial do traço latente e para a seleção do próximo item.

Esses testes têm sido utilizados em licenças profissionais, exames de admissão e de proficiência, como o *Test of English as a Foreign Language (TOEFL)*, que hoje é aplicado em mais de 130 países e utilizado até mesmo para admissões em programas de pós-graduação de grandes universidades americanas e europeias.

Além do *TOEFL*, nos Estados Unidos, o *Graduate Record Examination (GRE)* e o

*Armed Services Vocational Aptitude Test Battery (ASBAV)* também são aplicados como testes adaptativos. O primeiro também é utilizado na seleção de universitários, já o segundo serve para selecionar indivíduos para o serviço militar.

No Brasil, o CAT é utilizado na prova de Proficiência em Inglês da Universidade de Brasília (UnB) aplicado pelo Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (Cebbraspe) e o Computer-Aided Learning of English for Academic Purposes da Universidade de São Paulo.

O exame de Proficiência em Inglês é aplicado semestralmente e possibilita que alunos de graduação da UnB adquiram créditos referente à disciplina de Inglês Instrumental após bom desempenho em um teste em ambiente CAT.

Por se tratar de testes cuja escolha dos itens leva em consideração a estimativa parcial do traço latente do respondente, não há como se imaginar que essa adaptação seja possível sem a capacidade moderna de processamento dos computadores. Dessa forma, a escolha do item que será selecionado para o teste tem muito a ver com as respostas que já foram dadas para os itens anteriores do mesmo teste. A verossimilhança associada a essa etapa de seleção do item é dada por:

$$L(\theta|u_{1j}\dots u_{(k-1)j}) = \prod_{i=1}^{k-1} [P_i(\theta_j)]^{u_{ij}} [1 - P_i(\theta_j)]^{1-u_{ij}}. \quad (1.4)$$

Na abordagem bayesiana, uma priori do parâmetro do traço latente,  $g(\theta)$ , com a verossimilhança fornece uma distribuição a posteriori de  $\theta$ :

$$g(\theta|u_{1j}\dots u_{(k-1)j}) = \frac{L(\theta|u_{1j}\dots u_{(k-1)j})g(\theta)}{\int L(\theta|u_{1j}\dots u_{(k-1)j})g(\theta)d\theta}. \quad (1.5)$$

Normalmente, a densidade assumida é a uniforme ou uma estimativa empírica da distribuição do traço latente na população.

### 1.6.1 Estimativa do traço latente

Ao maximizar a verossimilhança em (1.4), em  $\theta$ , é possível obter o Estimador de Máxima Verossimilhança (EMV) para  $\theta$ , conforme a seguir:

$$\hat{\theta}_{u_{1j}\dots u_{(k-1)j}}^{EMV} = \operatorname{argmax}_{\theta} \{L(\theta|u_{1j}\dots u_{(k-1)j}) : \theta \in (-\infty, \infty)\}.$$

Warm (1989) propôs um estimador de máxima verossimilhança ponderado (EMVP) pela função  $w_{k-1}(\theta)$  dado por:

$$\hat{\theta}_{u_{1j}\dots u_{(k-1)j}}^{EMVP} = \operatorname{argmax}_{\theta} \{w_{k-1}(\theta)L(\theta|u_{1j}\dots u_{(k-1)j}) : \theta \in (-\infty, \infty)\},$$

em que  $w_{k-1}(\theta)$  é definido de tal forma que satisfaça:

$$\frac{\partial w_{k-1}(\theta)}{\partial \theta^2} = \frac{H_{k-1}(\theta)}{2I_{k-1}(\theta)},$$

com

$$H_{k-1}(\theta) = \sum_{i=1}^{k-1} \frac{[p'_{ij}(\theta)][p''_{ij}(\theta)]}{p_{ij}(\theta)[1 - p_{ij}(\theta)]},$$

$$p''_{ij}(\theta) = \frac{\partial^2 p_{ij}(\theta)}{\partial \theta^2}$$

e  $I_{k-1}(\theta)$  é a informação do teste após  $(k - 1)$  itens respondidos.

Outras alternativas de estimadores para  $\theta$  são baseadas na distribuição à posteriori em (1.5). Os mais utilizados são o MAP (*maximum a posteriori*), proposto na TRI por Lord (1986) e Mislevy (1986a); e o EAP (*expected a posteriori*), discutido amplamente por Bock e Mislevy (1982).

O estimador MAP de  $\theta$  é dado pela equação a seguir:

$$\hat{\theta}_{u_{1j}\dots u_{(k-1)j}}^{MAP} = \operatorname{argmax}_{\theta} \{g(\theta|u_{1j}\dots u_{(k-1)j}) : \theta \in (-\infty, \infty)\}.$$

Já o estimador EAP de  $\theta$ , baseado no valor esperado a posteriori, é dado por:

$$\hat{\theta}_{u_{1j}\dots u_{(k-1)j}}^{EAP} = \int \theta g(\theta|u_{1j}\dots u_{(k-1)j}) d\theta.$$

Uma importante medida para estimativa do  $\theta$  ao final do teste é a variância da distribuição à posteriori, que resume o grau de incerteza sobre o traço latente estimado. Essa variância é dada por:

$$\operatorname{Var}(\theta|u_{1j}\dots u_{(k-1)j}) = \int [\theta - E(\theta|u_{1j}\dots u_{(k-1)j})]^2 g(\theta|u_{1j}\dots u_{(k-1)j}) d\theta.$$

No modelo logístico de três parâmetros, o estimador de máxima verossimilhança pode apresentar problemas de estimação quando todas as respostas estão corretas ou todas

incorretas, por exemplo. Além disso, segundo Van der Linden e Pashley (2009), em testes adaptativos com banco de itens pequeno, as propriedades do EMV dependem de fatores como a distribuição de itens no banco e o critério de seleção dos itens.

Para a priori uniforme, o MAP possui propriedades semelhantes ao EMV. Já para outras prioris, deve-se tomar cuidado na busca do máximo global, pois a distribuição a posteriori pode ser multimodal e, dessa forma, a maximização pode retornar algum máximo local.

Para uma priori própria, o estimador EAP sempre existe e é de fácil obtenção, já que nenhum procedimento iterativo é exigido e um passo de método numérico para a integração já é suficiente.

Nos testes adaptativos, o traço latente é estimado em três momentos: a estimativa de um valor inicial para iniciar o processo de seleção dos itens; a estimativa, após cada item respondido, a fim de selecionar os itens que mais possuem informação na faixa estimada do traço latente; e a estimativa final, após conclusão do teste. Para cada um desses momentos, há uma ou mais metodologias adequadas.

No início do teste adaptativo, as chances de respostas totalmente corretas ou totalmente incorretas são grandes, e, conforme já visto, o estimador de máxima verossimilhança não é eficaz para estimar o traço latente nestes casos.

Van der Linden e Pashley (2009) trazem algumas soluções para essa etapa. Uma delas é estabelecer valores extremos na dificuldade para os itens aplicados e, após um número finito de itens, passar a utilizar o EMV. Outra alternativa é adiar a estimação do traço latente até que um número fixo de itens seja respondido. A terceira é a utilização de métodos bayesianos, não suscetíveis a essa limitação de estimação. A quarta solução seria, quando possível, a utilização de testes anteriormente aplicados para os mesmos respondentes.

Cada solução apontada tem suas desvantagens: as duas primeiras, por estabelecer valor arbitrário para a quantidade de itens que serão respondidos antes de iniciar a estimativa do traço latente, a solução bayesiana, por propor uma distribuição a priori que pode estar fora da realidade da estimativa, e a última solução, por exigir um cuidado no viés que pode ser gerado ao utilizar outro teste como referência para estimar o traço latente pela TRI.

Van der Linden e Pashley (2009) destacam, ainda, que esse problema só é relevante para testes curtos, com até 20 itens, pois os testes longos terão tempo de recuperar qualquer estimativa inicial insatisfatória.

Na etapa posterior, o método de estimativa deve ser aquele que proporcionará a convergência mais rápida. Não há nenhuma combinação de método de estimação e critério

de escolha do item que seja considerado excelente, mas algumas combinações são mais utilizadas. São o caso da utilização de EMV para estimar o traço latente e seleção do item pela máxima informação, ou a utilização de estimação bayesiana com a seleção do item baseada na distribuição a posteriori.

Para a estimativa final do traço latente, é preciso levar em conta que ela não irá guiar mais o processo de escolha do próximo item, mas irá transmitir ao examinado uma medida resumo sobre seu desempenho no teste. Entretanto, características do banco de itens e a escolha do método utilizado para seleção de itens influenciam essa etapa e precisam ser discutidos antes de uma definição do melhor método de estimação final do traço latente.

### 1.6.2 Critérios de Seleção do Item

Birnbaum (1968) propôs o uso da função de informação do teste para a seleção do item. A vantagem do uso dessa função é a facilidade de incorporar a informação de cada item selecionado. Sendo assim, o critério para a seleção do  $k$ -ésimo item é maximizar (1.3) em  $\theta = \hat{\theta}_{u_{1j}, \dots, u_{(k-1)j}}$ . De modo mais formal, a escolha do próximo item é feita por:

$$i_k = \underset{i}{\operatorname{argmax}} \{ I_{U_i}(\hat{\theta}_{u_{1j}, \dots, u_{(k-1)j}}) : i \in R_k \}, \quad (1.6)$$

sendo  $i_k$  o  $k$ -ésimo item selecionado para o teste e  $R_k$  o conjunto de itens no banco que ainda não foram utilizados.

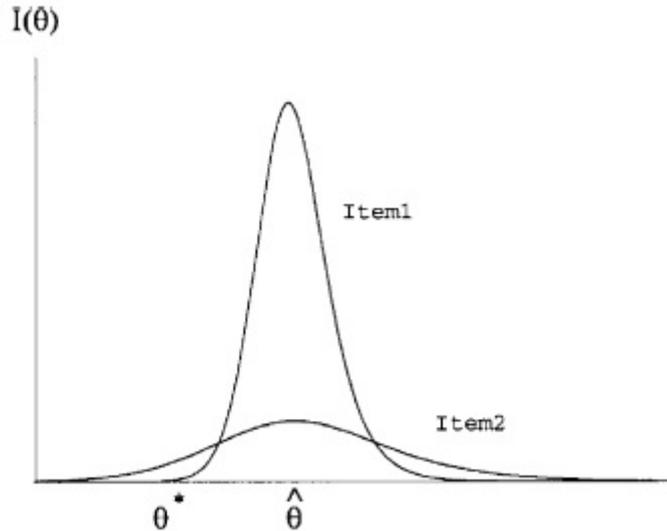
Chang e Ying (2009) mostram que a combinação do uso do EMV, para estimação de  $\theta$ , com o critério de seleção pela máxima informação, fornece um estimador de  $\theta$  que converge para o valor real com variância amostral próxima ao limitante mínimo fornecido pela informação de Fisher. Esse resultado só é válido para um banco de itens grande (ou infinito) com todos os possíveis valores de  $a_i$  entre 0 e  $\infty$  e valores de  $c_i$  limitado a 1. Além disso, no modelo logístico de três parâmetros, uma modificação na função de verossimilhança é necessária para evitar raízes múltiplas no processo de maximização.

Critérios de seleção do item em CAT devem considerar dois tipos de erros distintos: de estimação do traço latente já no ambiente de teste e na estimação dos parâmetros do item na fase de calibração antes da aplicação do teste.

Como o erro de estimação do traço latente costuma ser alto no início do teste, ignorá-lo pode conduzir a condições enganosas de otimalidade de  $\theta$ . Na Figura 1.2, a curva característica de dois itens mostra que na estimativa atual,  $\hat{\theta}$ , o item com máxima informação é

o item 1, mas no verdadeiro valor do traço latente,  $\theta^*$ , o item com máxima informação é o item 2.

Figura 1.2: Seleção do item no nível de traço latente impreciso.



Fonte: Van der Linden e Pashley (2009)

Os erros de estimação dos parâmetros dos itens podem ser ignorados quando a amostra utilizada para a calibração é grande. Além disso, após uma boa calibração, somente itens adequados permanecem no teste. Entretanto, quanto maior for a amostra de calibragem, menor é a vida útil de cada item, pois eles seriam expostos para uma quantidade maior de respondentes, gerando maiores custos de elaboração da avaliação como um todo.

Chang e Ying (1996) propuseram a utilização da informação de Kullback-Leibler ao invés da informação de Fisher na escolha do item, a fim de amenizar problemas com erros de estimativa no início do teste. Em geral, a informação de Kullback-Leibler é utilizada para medir distância entre duas distribuições, sendo que, quanto maior for a informação, mais fácil é discriminar entre duas distribuições.

Para o modelo logístico de três parâmetros, a medida de Kullback-Leibler no  $k$ -ésimo item com o verdadeiro valor do traço latente,  $\theta^*$ , e a atual estimativa do traço latente,  $\hat{\theta}_{k-1}$ , é dada por:

$$K_{i_k}(\hat{\theta}_{k-1}, \theta^*) = E \left[ \log \frac{L(\theta_0 | U_{i_k})}{L(\hat{\theta}_{k-1} | U_{i_k})} \right]. \quad (1.7)$$

Outra forma de representar a informação de Kullback-Leibler é:

$$K_{i_k}(\hat{\theta}_{k-1}, \theta^*) = p_{i_k}(\theta^*) \log \frac{p_{i_k}(\theta^*)}{p_{i_k}(\hat{\theta}_{k-1})} + [1 - p_{i_k}(\theta^*)] \log \frac{1 - p_{i_k}(\theta^*)}{1 - p_{i_k}(\hat{\theta}_{k-1})}.$$

Conforme já foi dito, a informação de Kullback-Leibler mede a distância da estimativa atual para o valor verdadeiro do traço latente. Porém, este valor verdadeiro é desconhecido e, por isso, Chang e Ying (1996) propuseram trocar o apresentado em (1.7) pela integral sobre todo intervalo em torno da última estimativa do traço latente,  $[\hat{\theta}_{k-1} - \delta_k, \hat{\theta}_{k-1} + \delta_k]$ , sendo  $\delta_k$  uma função decrescente da ordem de seleção do item no teste. O  $k$ -ésimo item no teste é selecionado de acordo com:

$$i_k = \operatorname{argmax}_i \left\{ \int_{\hat{\theta}_{k-1} - \delta_k}^{\hat{\theta}_{k-1} + \delta_k} K_i(\hat{\theta}_{k-1}, \theta) d\theta : i \in R_k \right\}.$$

Veerkamp e Berger (1997) propuseram uma integração semelhante, mas para todos os possíveis valores de  $\theta$  e utilizando a informação de Fisher, mas a metodologia pode facilmente ser adaptada para a medida de Kullback-Leibler.

A proposta consiste numa função de verossimilhança ponderada pela informação de Fisher e seleciona o  $k$ -ésimo item por:

$$i_k = \operatorname{argmax}_i \left\{ \int_{-\infty}^{\infty} L(\theta | u_{1j}, \dots, u_{(k-1)j}) I_{i_k}(\theta) d\theta : i \in R_k \right\}. \quad (1.8)$$

Se a estimação de  $\theta$  é feita pelo estimador de máxima verossimilhança, então o critério de Veerkamp e Berger dará mais peso aos valores de  $\theta$  mais próximos ao traço latente atual. No início do teste, valores mais distantes de  $\hat{\theta}_{k-1}$  recebem pesos substanciais, mas à medida que o teste se aproxima do fim, o peso se concentra em valores próximos de  $\hat{\theta}_{k-1}$ .

As abordagens bayesianas envolvem o uso da distribuição a posteriori de  $\theta$ . Como a distribuição a posteriori é uma combinação da verossimilhança e a distribuição a priori, a única diferença com os critérios apresentados anteriormente é a suposição da priori. Geralmente considera-se priori pouco informativa para evitar erros de estimação devido a escolha da priori, a menos que haja alguma informação disponível do respondente.

Um critério de informação a posteriori ponderado pode ser definido por:

$$i_k = \operatorname{argmax}_i \left\{ \int I_{U_i}(\theta) g(\theta | u_{1j}, \dots, u_{(k-1)j}) d\theta : i \in R_k \right\}. \quad (1.9)$$

Esse critério também é baseado na informação esperada de Fisher e, apesar da diferença entre a informação esperada e a observada para modelos logísticos de três parâmetros, a informação observada também pode ser utilizada. Ademais, o critério também admite a utilização da informação de Kullback-Leibler.

Outros critérios de seleção do item com abordagem bayesiana são apresentados por Van der Linden e Pashley (2009).

Os critérios apresentados até então têm motivação estatística. Um critério que, além da motivação estatística, também considera a praticidade é o método de multiestágio  $\alpha$ -estratificado proposto por Chang e Ying (1999). A princípio, o método foi proposto para reduzir o efeito do erro de estimação do traço latente na seleção do item. Como foi mostrado na Figura 1.2, se os erros são grandes, um item com menor discriminação pode ser mais eficiente para um intervalo maior de  $\theta$  que outro com maior discriminação.

A proposta consiste em estratificar o banco de itens de acordo com os valores dos parâmetros de discriminação e restringir a seleção do item ao estrato com valores crescentes durante o teste. Em cada estrato, itens são selecionados de acordo com o critério de distância mínima entre o valor do parâmetro de dificuldade e a estimativa atual do traço latente. Chang e Ying (2008) mostraram por que a seleção precoce de itens altamente discriminantes após algumas respostas iniciais incorretas é prejudicial para a estimação de  $\theta$ . Este procedimento também resolve o problema de exposição desigual dos itens, pois itens com menor discriminação têm uma chance igual de serem escolhidos.

A questão sobre a melhor combinação de critério de seleção do item e estimação do traço latente só é analisada, atualmente, sob aspectos assintóticos.

Um resultado bem conhecido da estatística bayesiana é que, para  $k \rightarrow \infty$ , a distribuição a posteriori  $g(\theta | u_{1j}, \dots, u_{(k-1)j})$  converge para degenerações no verdadeiro valor de  $\theta$ . Conclui-se, então, que toda estimação do traço latente baseada na posteriori e procedimentos de seleção do item vistos aqui produzem resultados assintóticos idênticos. Também foi mostrado que a combinação do EMV com o critério de máxima informação fornece boas estimativas para  $\theta$ .

Van der Linden e Pashley (2009) destacam que os testes adaptativos, por terem menos

itens, necessitam de simulações para comparação de combinação de métodos. Estudos relevantes nesse aspecto foram desenvolvidos por Chang e Ying (1999), van der Linden (1998), Veerkamp e Berger (1997), Wang et al. (1999), Wang e Vispoel (1998), Weiss (1982), Weiss e McBride (1984) e Warm (1989).

Propostas mais recentes de critério de seleção do item incluem adaptações à estratificação proposta por Chang e Ying (1999), tal como o trabalho de Chang e van der Linden (2003); utilização do tempo de resposta para escolha do item, Fan et al. (2012) e Choe et al. (2017); e critérios para testes específicos, como a proposta de Kaplan et al. (2015) para diagnóstico cognitivo.

## 1.7 Regressão Isotônica

A regressão isotônica, que tornou-se mais conhecida após trabalho de Barlow et al. (1972), refere-se a uma regressão não paramétrica que minimiza a soma dos quadrados de uma função não decrescente. Diferencia-se da regressão usualmente utilizada por se valer da informação de uma relação de ordem da variável resposta.

Seja  $X$  um conjunto finito  $x_1, \dots, x_n$  com relação de ordem estabelecida tal que  $x_1 < x_2 < \dots < x_n$ . Uma função real  $f$  em  $X$  é isotônica se  $x, y \in X$  e  $x < y$  implica que  $f(x) \leq f(y)$ , ou seja, é uma função não decrescente.

Se  $g$  é uma função em  $X$  e  $w$  uma função positiva em  $X$ , então uma função isotônica  $g^*$  em  $X$  é uma regressão isotônica de  $g$ , com peso  $w$  e com respeito a relação de ordem  $x_1 < x_2 < \dots < x_n$ , se ela minimiza na classe das funções isotônicas  $f$  em  $X$ , a soma

$$\sum_{i=1}^n [g(x_i) - f(x_i)]^2 w(x_i).$$

Quando os parâmetros de interesse apresentam algum tipo de ordenação, espera-se que a regressão isotônica seja mais adequada, justamente por considerar essa relação de ordem. É o caso da relação entre a dificuldade do item e a probabilidade de acerto, pois quanto mais difícil é o item, menos acertos espera-se do respondente, mantido o parâmetro  $a$  constante, tendo, assim, uma relação de ordem decrescente, podendo torna-se crescente simplesmente com uma inversão do eixo relativo a dificuldade do item.

Uma interpretação gráfica esclarecedora da regressão isotônica pode ser obtida ao plotar

as seguintes somas cumulativas:

$$G_k = \sum_{i=1}^k g(x_i)w(x_i),$$

contra

$$W_k = \sum_{i=1}^k w(x_i), \quad k = 1, 2, 3, \dots, n.$$

Esse gráfico nada mais é do que uma plotagem dos pontos  $P_k = (W_k, G_k)$  no plano cartesiano. Esses pontos constituem o diagrama da soma acumulada (DSA) de uma função dada  $g$  com peso  $w$ . A inclinação do segmento que liga o ponto  $P_{k-1}$  a  $P_k$  é dada por:

$$\frac{G_k - G_{k-1}}{W_k - W_{k-1}} = g(x_k).$$

Já a inclinação do segmento que liga  $P_{i-1}$  a  $P_k$ , para  $i \leq k$ , representa a média ponderada

$$Av\{x_i, x_{i+1}, \dots, x_k\} = \frac{\sum_{r=i}^k g(x_r)w(x_r)}{\sum_{r=i}^k w(x_r)}.$$

Será mostrado que a regressão isotônica de  $g$  é dada pela inclinação do minorante convexo máximo (MCM) do DSA.

A função minorante convexa máxima (MCM) do DSA no intervalo  $[0, W_n]$ ,  $G^*$ , é o gráfico do supremo de todas as funções convexas que estão abaixo do DSA. Trata-se de uma linha que intercepta o DSA em alguns pontos, mas que nunca localiza-se acima dele.

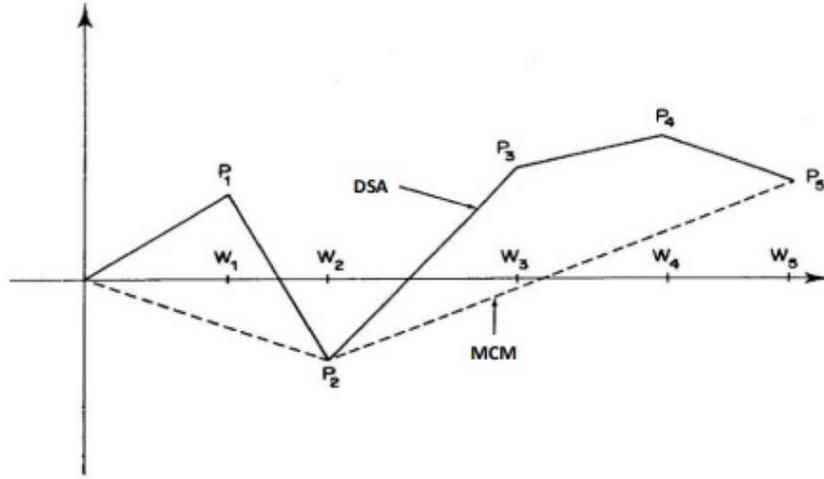
A Figura 1.3 ilustra o DSA e a função MCM de um conjunto de dados hipotéticos. Como  $G^*$  é convexo, por definição, no intervalo  $[0, W_n]$ , então também será diferenciável em cada um dos pontos  $W_i$ . Os pontos da MCM são representados por  $P_k^* = (W_k, G_k^*)$  com

$$G_k^* = \sum_{i=1}^k g^*(x_i)w(x_i).$$

Desta forma, o valor da regressão isotônica  $g^*$  no ponto  $x_k$  é a inclinação da MCM no ponto  $P_k^* = (W_k, G_k^*)$ .

Se  $P_k^*$  for um ponto não diferenciável do gráfico,  $g^*(x_k)$  será a inclinação do seguimento prolongado à esquerda. Se para algum  $i$  a MCM em  $W_i$  for posta estritamente abaixo do DSA, então a inclinação da MCM à esquerda e à direita de  $W_i$  são as mesmas.

Figura 1.3: Representação gráfica do Diagrama de Soma Acumulada (DSA) e a função Minorante Convexa Máxima (MCM)



Fonte: Schlemper (2014)

Algumas propriedades do MCM serão expressas analiticamente. A primeira delas é que o DSA e o MCM coincide em  $P_n^*$ , ou seja,

$$G_n^* = G_n. \quad (1.10)$$

A segunda propriedade é que, se para algum índice  $i$  o MCM em  $P_{i-1}^*$  mantém-se estritamente abaixo do DSA em  $P_{i-1}$ , então a inclinação do MCM chegando em  $P_{i-1}^*$  e saindo do mesmo são iguais. Sendo assim,

$$G_{i-1}^* < G_{i-1} \Rightarrow g_i^* - g_{i-1}^* = 0, \quad i = 1, 2, \dots, n. \quad (1.11)$$

Finalmente, se  $P_r = P_r^*$ ,  $P_s = P_s^*$ , se  $P_r P_s = P_r^* P_s^*$  é um vértice do MCM e  $P_s P_t = P_s^* P_t^*$  é o vértice adjacente a direita, e se  $r < q < s$ , então a inclinação de  $P_q P_s$  é menor que a inclinação de  $P_q^* P_s^*$ . Se  $s < q < t$  então a inclinação de  $P_s P_q$  é maior que a inclinação de  $P_s^* P_q^*$ .

Essas propriedades auxiliam na prova de que  $g^*$  é uma regressão isotônica de  $g$  ponderado por  $w$ .

**Teorema 1:** *Se  $X$  tem ordenação simples, a inclinação  $g^*$  do MCM fornece a regressão*

isotônica de  $g$ . De fato, se  $f$  é isotônica em  $X$  então

$$\sum_x [g(x) - f(x)]^2 w(x) \geq \sum_x [g(x) - g^*(x)]^2 w(x) + \sum_x [g^*(x) - f(x)]^2 w(x). \quad (1.12)$$

A regressão isotônica é única.

**Prova:** A diferença entre a expressão da esquerda e da direita de (1.12) é

$$2 \sum_x [g(x) - g^*(x)][g^*(x) - f(x)]w(x).$$

Assim, basta mostrar que

$$\sum_x [g(x) - g^*(x)][g^*(x) - f(x)]w(x) \geq 0$$

para toda regressão isotônica  $f$  em  $X$ . A fórmula da soma parcial de Abel fornece

$$\begin{aligned} & \sum_{i=1}^n [g(x_i) - g^*(x_i)][g^*(x_i) - f(x_i)]w(x_i) = \\ & \sum_{i=1}^n \{[f(x_i) - f(x_{i-1})] - [g^*(x_i) - g^*(x_{i-1})]\} \\ & \times [G_{i-1} - G_{i-1}^*]w(x_i) + [g^*(x_n) - f(x_n)][G_n - G_n^*], \end{aligned}$$

em que  $x_0 = f(x_0) = g^*(x_0) = G_0 = G_0^* = 0$ . O último termo na direita é 0 por (1.10). Pela propriedade apresentada em (1.11),  $[g^*(x_i) - g^*(x_{i-1})][G_{i-1} - G_{i-1}^*] = 0$ ,  $i = 1, 2, \dots, n$ . Visto que  $G_{i-1}^* \leq G_{i-1}$  e  $f(x_{i-1}) \leq f(x_i)$ ,  $i = 1, 2, \dots, n$ , pela última propriedade. Para verificar a unicidade da regressão isotônica, suponha  $g_1$  como outra função isotônica em  $X$  e que

$$\sum_x [g(x) - g_1(x)]^2 w(x) \leq \sum_x [g(x) - f(x)]^2 w(x)$$

para toda função isotônica  $f$  em  $X$ . Então sendo  $g_1$  e  $g^*$  isotônicas,

$$\sum_x [g(x) - g_1(x)]^2 w(x) = \sum_x [g(x) - g^*(x)]^2 w(x).$$

Porém, de (1.12), sendo  $g_1$  isotônica,

$$\sum_x [g(x) - g_1(x)]^2 w(x) \geq \sum_x [g(x) - g^*(x)]^2 w(x) + \sum_x [g^*(x) - g_1(x)]^2 w(x),$$

que implica que

$$\sum_x [g^*(x) - g_1(x)]^2 w(x) = 0,$$

consequentemente  $g^*(x) = g_1(x)$  para  $x \in X$ .

### 1.7.1 Algoritmo PAV

Fica claro que se, para algum  $i$ ,  $g(x_{i-1}) > g(x_i)$ , então o gráfico do MCM entre os pontos  $P_{i-2}^*$  e  $P_i^*$  é um segmento de linha reta. Então o gráfico do DSA pode ser alterado pela conexão de  $P_{i-2}^*$  e  $P_i^*$  por um segmento de linha reta, sem alteração no MCM. Este método de sucessivas aproximações do MCM pode ser descrito algebricamente como algoritmo *pool-adjacent-violators (PAV)*, proposto por Ayer et al. (1955).

Resumidamente, o algoritmo PAV é dado pelos seguintes passos. Considerando  $X_i$  a variável explicativa e  $Y_i = f(x_i)$  a variável explicada:

1. Ordene os pares  $(X_i, Y_i)$  de acordo com os valores de  $X_i$ , obtendo  $(X_{(i)}, Y_{(i)})$ , em que  $(i)$  é a  $i$ -ésima observação após a ordenação.
2. Começando por  $Y_{(1)}$ , mova para a direita e pare se o par  $(Y_{(i)}, Y_{(i+1)})$  violar a restrição de monotonicidade, ou seja, se  $Y_{(i)} > Y_{(i+1)}$ . Se a restrição for violada, substitua os valores de  $Y_{(i)}$  e  $Y_{(i+1)}$  pela média de ambos, de tal forma que  $Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i)} + Y_{(i+1)})/2$ .
3. Confira se  $Y_{(i-1)} \leq Y_{(i)}^*$ . Se não, substitua os valores de  $Y_{(i-1)}$ ,  $Y_{(i)}$  e  $Y_{(i+1)}$  pela média aritmética dos três, de tal forma que  $Y_{(i-1)}^* = Y_{(i)}^* = Y_{(i+1)}^* = (Y_{(i-1)} + Y_{(i)} + Y_{(i+1)})/3$ .

Seguindo esses passos até obter a restrição de ordem necessária, o algoritmo fornecerá a regressão isotônica.

Härdle (1990) destaca que, se os dados seguem a restrição de ordem, então o algoritmo PAV irá simplesmente reproduzir os dados, e se a restrição for violada para algum  $Y_i$ , então uma função escada será produzida e degraus são obtidos no intervalo de violação. Essa função escada não é suavizada, embora seja não decrescente. Além disso, na presença de *outliers* ou valores discrepantes, o algoritmo PAV fornece uma função com uma longa linha reta na região desses valores.

Uma solução proposta por Lee (2007) seria aplicar o algoritmo PAV para garantir a restrição e, depois, suavizar a regressão.

## 1.7.2 Método de Nadaraya-Watson

Um dos métodos conhecidos e, talvez, o mais utilizado para suavizar uma função de regressão é o estimador via kernel de Nadaraya-Watson,  $\hat{m}(x)$ , que é dado pela seguinte equação:

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)Y_i}{\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)},$$

em que  $K$  é uma função kernel simétrica, não negativa e contínua que determina a forma dos pesos kernel; e  $h$  indica a largura de janela, que determina o tamanho dos pesos.

O valor de  $h$ , também conhecido como parâmetro de suavização, é importante, pois além de indicar o tamanho dos pesos, serve para controlar a importância da suavização na estimação da regressão. Este parâmetro vai determinar o quão longe se permite que a suavização esteja dos degraus estimados no algoritmo PAV.

Um parâmetro de suavização pequeno produzirá estimadores mais severos, que dependem fortemente dos dados próximos de  $x$ , enquanto valores altos providenciam valores mais suaves.

Na prática a escolha do parâmetro é feita de forma a balancear o viés e a variância da estimação, componentes do erro quadrático médio. Valores mais baixos para  $h$  produzem viés pequenos, mas variância alta e valores mais altos produzem variância pequena e viés alto. Adota-se, portanto, um valor para o parâmetro de suavização que minimize o erro quadrático médio.

## 1.8 Considerações

A regressão isotônica já tem sido utilizada em TRI para estimar a curva característica do item (Schlemper, 2014) e para imputação de dados faltantes (Pereira, 2014).

Entretanto, o interesse neste trabalho é encontrar a regressão isotônica da variável resposta indicadora de acerto de uma questão, com respeito ao seu nível de dificuldade, ou  $b_i$ . Obviamente, a função de acerto do item com respeito a sua dificuldade será decrescente, pois, mantendo o parâmetro de discriminação constante, quanto mais difícil é o item, menor é a probabilidade de acerto.

Para contornar esse problema, a dificuldade do item será considerada em seus valores decrescentes, de tal forma que a função de probabilidade de acerto dada a dificuldade do

item seja vista da direita para esquerda, tornando-se assim numa função não decrescente e possibilitando o uso de regressão isotônica.

Para a proposta de novo critério de seleção de itens, será feita a estimativa do traço latente por meio de EAP e para escolha do item seguinte será suavizada, via Nadaraya-Watson, uma função de regressão isotônica com variável resposta a indicadora de acerto do item e variável explicativa a dificuldade dos itens a serem selecionados. Uma alternativa à suavização por Nadaraya-Watson também será apresentada, que consiste em uma interpolação linear no segmento que contém o valor desejado para a variável explicada.

Para a aplicação do método de Nadaraya-Watson, a informação total dos itens já respondidos para o valor corrente da estimativa do traço latente será utilizada como ponderação.

Após a suavização, será identificado o item com maior informação, dentre aqueles ainda não utilizados, considerando como nível de traço latente o parâmetro de dificuldade cuja probabilidade de acerto, estimada pela regressão suavizada, é igual a 0,6. Esse valor foi escolhido por representar nas simulações o valor de  $(1 + c)/2$ , que é o valor onde é calculado a discriminação e a dificuldade do item. O método será melhor detalhado no capítulo seguinte.

## Capítulo 2

# Metodologia Proposta

A metodologia que se propõe diz respeito a fase de escolha do item. Nos estudos de simulação será verificada a combinação do critério de escolha do item baseado na regressão isotônica junto com a estimação do traço latente via EAP.

O pacote *mirtCAT* do *software R* (R Core Team), desenvolvido por Chalmers (2016), já permite que seja simulada a aplicação de um teste adaptativo computadorizado utilizando diversos critérios para seleção do item, como a máxima informação, a mínima variância esperada à posteriori, informação de máxima verossimilhança ponderada e critérios utilizando a integração de Kullback-Leibler.

Os critérios de convergência e fim da aplicação do teste também são os mais distintos e envolvem o erro padrão, diferença dos  $\theta$ 's e número de itens pré-estabelecidos. O mais utilizado é o erro padrão, que será utilizado nas simulações com uma restrição de número máximo de itens.

Para manter as definições padrões na utilização de cada critério de seleção do item, será desenvolvido *script* da nova metodologia proposta, bem como das metodologias já implementadas no pacote *mirtCAT*.

Para a formação dos bancos de itens foram simulados os parâmetros de 100.000 itens, de tal forma que  $a_i \sim \mathcal{U}(0.6, 2.5)$ ,  $b_i \sim \mathcal{N}(0, 2.5)$  e  $c_i = 0.2$  para todo  $i = 1, 2, \dots, 100.000$ . O valor do parâmetro de escala será considerado  $D = 1,702$ , que aproxima o modelo de uma ogiva normal. O valor mínimo de  $a_i$  é para permitir que o produto  $Da_i$  seja maior do que 1, o valor máximo de  $a_i$  e o desvio padrão de  $b_i$  é para ter itens suficientes com  $b > 3$  e  $b < 3$ , permitindo uma boa estimativa para esses níveis de traço latente nas simulações.

Os itens serão estratificado em 100 bancos de 1.000 itens cada, para que se possa ter

uma média dos valores estimados em cada ponto do traço latente.

A estimativa inicial do traço latente será 0, que corresponde a média da métrica do traço latente. Já a distribuição a priori de  $\theta$  será uma  $\mathcal{N}(0, 1)$ .

A quantidade de itens poderia ser maior para a avaliação da nova metodologia. Entretanto, optou-se por limitar a um banco de itens de tamanho coerente com os tamanhos dos bancos de itens reais na área de avaliação educacional.

No início do teste duas respostas fictícias são atribuídas a itens com parâmetro de dificuldade extremos, garantindo, assim, que o processo de estimação e escolha dos próximos itens seja bem sucedido, com uma resposta correta e outra incorreta, respectivamente. Esses itens não são utilizados para a estimação de  $\theta$  e servem, tão somente, para permitir um melhor ajuste da regressão isotônica. Nas simulações deste trabalho foram escolhidos itens com parâmetro de dificuldade igual a  $b_{min} - 1$  e  $b_{max} + 1$ , sendo  $b_{min}$  o menor  $b$  dentre os já selecionados e  $b_{max}$  o maior.

Os próximos dois itens serão os dois primeiros que comporão verdadeiramente o teste. Eles obedecerão as seguintes restrições:

$$a_i \leq 1, i=1,2 \text{ e}$$

$$-0.5 \leq b_1 \leq 0 \text{ e } 0 \leq b_2 \leq 0.5$$

A restrição no parâmetro de discriminação permite que o teste incremente informação relevante para qualquer nível do traço latente do respondente, evitando o problema descrito na Figura 1.2. Já a restrição do parâmetro de dificuldade serve para direcionar a escolha dos próximos itens na escala do traço latente. Caso o respondente erre os dois primeiros itens, começa-se a seleção de itens com  $b_i < -0.5$ , caso acerte o item mais fácil, seleciona-se itens entre  $-0.5$  e  $0.5$  e, por fim, se a resposta aos dois primeiros itens for correta, segue-se para seleção de itens com  $b_i > 0.5$ .

A resposta simulada é resultado de um experimento aleatório de Bernoulli com parâmetro  $p_{ij}$ , ou seja, a probabilidade de acerto do item, conforme modelo descrito na Equação (1.1) com  $\theta$  igual ao nível do traço latente real na simulação. Se o experimento for bem sucedido, considera-se o acerto no item  $i$ , se não, considera-se o erro no mesmo item.

A partir daí, é aplicado o uso da regressão isotônica na escolha do próximo item que será selecionado. Entretanto, o primeiro item sem a restrição do parâmetro de dificuldade, permanece sob a restrição de  $a_i < 1$ .

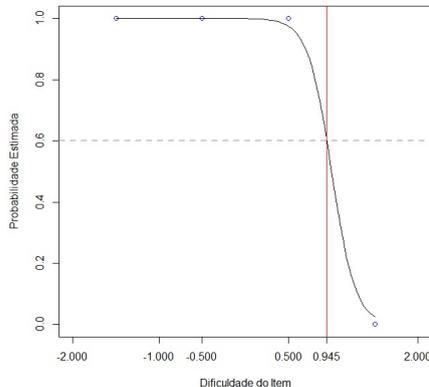
Considerando, como exemplo, um respondente que tenha acertado os dois primeiros itens, o vetor de dificuldade dos itens é  $\mathbf{b} = (-1.5, 1.5, -0.5, 0.5)$  e o vetor de resposta dos itens é  $\mathbf{Y} = (1, 0, 1, 1)$ . Após ordenação em  $\mathbf{b}$ :

$$\mathbf{b}_{ord} = (-1.5, -0.5, 0.5, 1.5) \text{ e } \mathbf{Y}_{ord} = (1, 1, 1, 0)$$

Como  $\mathbf{Y}_{ord}$  é não crescente em  $\mathbf{b}_{ord}$ , utiliza-se a regressão isotônica com inversão no eixo  $x$  para a escolha do item. A Figura 2.1 representa a regressão obtida ponderada pela informação de Fisher calculada com a estimativa de  $\theta$  corrente e, nela, é possível identificar o parâmetro de dificuldade cuja imagem seja 0.6, ou seja, a probabilidade estimada pela regressão é totalmente aleatória.

No exemplo ilustrado, o valor de  $\theta$  escolhido no critério da regressão isotônica suavizada por Nadaraya-Watson é próximo a 0.945. Esse valor que será utilizado para o cálculo da Informação de Fisher e posterior seleção do item com máxima informação.

Figura 2.1: Escolha do próximo item com uso da Regressão Antitônica suavizada por Nadaraya-Watson ( $h=1$ )



O pacote no  $R$  que tem implementado a regressão isotônica é o *fdrtool*, desenvolvido por Klaus e Strimmer (2015), mais especificamente a função *monoreg*. Para suavização da regressão utilizando Nadaraya-Watson via kernel, utiliza-se a função *ksmooth*. Neste caso, utilizou-se a largura da janela,  $h = 1$ , e como função kernel a densidade da Normal no ponto de  $\mathbf{b}_w - b_{med}$ , sendo  $\mathbf{b}_w$  os parâmetros de dificuldade dos itens respondidos ordenados e  $b_{med}$  o parâmetro de dificuldade meta para a escolha do próximo item, ou seja, aquele cuja imagem seja próxima da probabilidade de acerto aleatório.

Além da proposta de suavização via kernel, foi simulado também uma interpolação linear no segmento que contém a imagem 0.6, para encontrar a imagem do  $b_{med}$  desejada.

Ambos os critérios serão apresentados nos resultados da simulação.

Para encontrar a regressão isotônica é necessário ter uma estimativa parcial do traço latente do respondente. A estimativa do traço latente será feita pelo método da quadratura gaussiana, proposto por Gray (2001), para o estimador EAP dado pela seguinte expressão:

$$\begin{aligned}\hat{\theta}_j^{EAP} &= \frac{\int_{\mathbb{R}} \theta_j L(\theta_j; u_1, \dots, u_{k-1}) g(\theta_j) d\theta_j}{\int_{\mathbb{R}} L(\theta_j; u_1, \dots, u_{k-1}) g(\theta_j) d\theta_j} \\ &\approx \frac{\sum_{t=1}^q \bar{\theta}_t L(\bar{\theta}_t; u_1, \dots, u_{k-1}) A_t}{\sum_{t=1}^q L(\bar{\theta}_t; u_1, \dots, u_{k-1}) A_t},\end{aligned}$$

em que  $\bar{\theta}_t$  representa os pontos de quadratura e  $A_t$  o peso associado a  $\bar{\theta}_t$ , definidos conforme metodologia descrita em Gray (2001) e melhor explicada no apêndice.

Ao final de cada resposta do item escolhido será calculado o desvio à posteriori,  $psd_j$ , de  $\hat{\theta}_j$  e se for maior do que o critério mínimo estabelecido, repete-se todo o processo de escolha do item; se não, finaliza-se o algoritmo com o último  $\hat{\theta}_j$  estimado. O desvio a posteriori associado ao EAP foi utilizado, também calculado pelo método de quadratura gaussiana, de tal forma que:

$$psd_j = \sqrt{[\hat{\theta}_j^2] - [\hat{\theta}_j]^2},$$

sendo  $[\hat{\theta}_j]^2$  o quadrado de  $\hat{\theta}_j^{EAP}$  e  $[\hat{\theta}_j^2]$  igual a:

$$\begin{aligned}\hat{\theta}_j^2{}^{EAP} &= \frac{\int_{\mathbb{R}} \theta_j^2 L(\theta_j; u_1, \dots, u_{k-1}) g(\theta_j) d\theta_j}{\int_{\mathbb{R}} L(\theta_j; u_1, \dots, u_{k-1}) g(\theta_j) d\theta_j} \\ &\approx \frac{\sum_{t=1}^q \bar{\theta}_t^2 L(\bar{\theta}_t; u_1, \dots, u_{k-1}) A_t}{\sum_{t=1}^q L(\bar{\theta}_t; u_1, \dots, u_{k-1}) A_t}.\end{aligned}$$

## 2.1 Simulações

Seguindo os exemplos de estudos de simulações apresentados nos trabalhos de Labarrere et al. (2011) e Costa (2009), as simulações serão separadas em duas partes, uma com banco de itens simulados e outra com um banco de itens da prova de Proficiência em Inglês da UnB.

No primeiro estudo, serão simulados 100 respondentes para cada valor de traço latente entre -3 e 3, variando 0.2 de um para outro, e dois valores distintos para o critério de parada, que é o desvio-padrão a posteriori: 0.2 e 0.1. Serão simulados, portanto, 3.100 respondentes para cada nível de erro padrão, sendo que serão 100 respostas para cada nível de  $\theta$ .

Para avaliar a qualidade da utilização do novo critério, será analisado o número de itens necessários para obter a convergência, o viés do traço latente estimada e o tempo computacional.

## Desvio Padrão a Posteriori = 0.2

Utilizando como critério de parada o desvio padrão a posteriori = 0.2, obteve-se convergência com poucos itens para todos os métodos. Importante destacar que houve uma restrição de 50 itens respondidos, ou seja, ou atingia-se o critério de parada ou respondia-se, no máximo, 50 itens. Na prática, dificilmente um teste será composto por tantos itens, mas como o objetivo aqui é comparar os critérios de escolha do item por simulações, então estabelece-se um número alto de itens máximos respondidos.

Os critérios comparados são os dois propostos neste trabalho, uso da regressão isotônica suavizada por Nadaraya-Watson e interpolação linear, mais quatro métodos já apresentados aqui, que são os critérios da máxima informação, proposto em (1.6); do produto da verossimilhança pela informação de Fisher, proposto em (1.8); produto da posteriori pela informação de Fisher, proposto em (1.9); e a adaptação utilizando Kullback-Leibler (1.7).

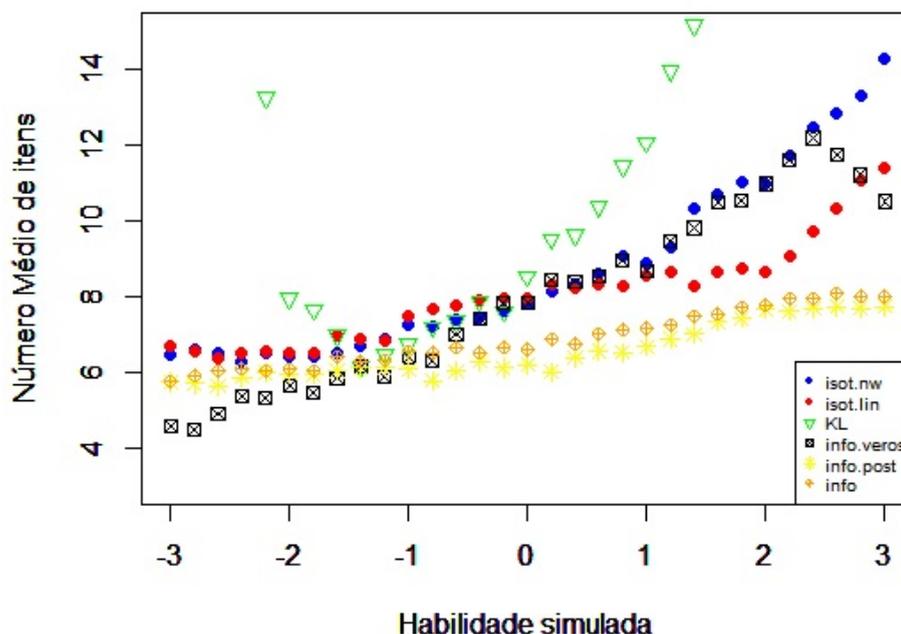
Tabela 2.1: Tabela resumo dos critérios comparados

Nome do Critério	Sigla nas tabelas
Regressão suavizada por Nadaraya-Watson	RINW
Regressão com Interpolação Linear	RIIL
Máxima Informação	MI
Verossimilhança ponderada pela Informação	VPI
Posteriori ponderada pela Informação	PPI
Máxima Informação de Kullback-Leibler	KL

A Figura 2.2 mostra a quantidade de itens médios utilizados na simulação de cada critério para cada ponto de traço latente simulado. Nela, é possível notar que a adaptação da informação de Kullback-Leibler (verde) não apresenta bons resultados para  $\theta$ 's fora do intervalo (-2;0).

Todos os métodos apresentam um aumento médio no número de itens utilizados para valores de  $\theta$  positivos, quando comparados com os negativos, principalmente para os dois critérios propostos por esse trabalho (vermelho e azul) e para o produto da verossimilhança com a informação de Fisher (preto). Este último, em particular, parece ser o melhor critério para valores extremamente negativos de  $\theta$  e, excluindo Kullback-Leibler e juntamente com

Figura 2.2: Quantidade média de itens utilizados em cada critério de escolha - Desvio Padrão a Posteriori = 0.2



a isotônica suavizada por Nadaraya-Watson, o pior critério para valores extremamente positivos de  $\theta$ .

Dentre os dois critérios propostos, o que demonstrou melhor desempenho, quando avaliado somente a quantidade média de itens utilizados, foi a regressão isotônica com interpolação linear (vermelho) para os valores de  $\theta > 0$ .

Já dentre todos os critérios, o que aparenta ter os melhores resultados é o critério que utiliza o produto da posteriori com a informação de Fisher (amarelo). Embora, no intervalo  $(-3;-1)$ , o critério do produto da verossimilhança com a informação de Fisher apresente melhores resultados.

A Figura 2.3 apresenta o vício para  $\theta$ 's positivos, que, geralmente, é o maior interesse em um CAT. Nota-se que todos os critérios possuem vícios semelhantes até  $\theta = 2.2$ , quando o critério do produto da verossimilhança com a informação de Fisher (preto) começa a se destacar positivamente, juntamente com os dois critérios propostos nesta dissertação.

A Figura 2.3 deve ser interpretada como um *plot* da habilidade simulada x habilidade estimada. A linha reta é onde os pontos deveriam estar se não houvesse nenhum vício. Quanto mais perto o ponto estiver da reta, mais acurado é o critério de escolha do item correspondente.

A Figura 2.4 mostra a mesma informação para os valores negativos de  $\theta$ . Nela, destaca-

Figura 2.3: Habilidade simulada x Habilidade estimada - Valores de  $\theta$  positivo - Desvio Padrão a Posteriori = 0.2

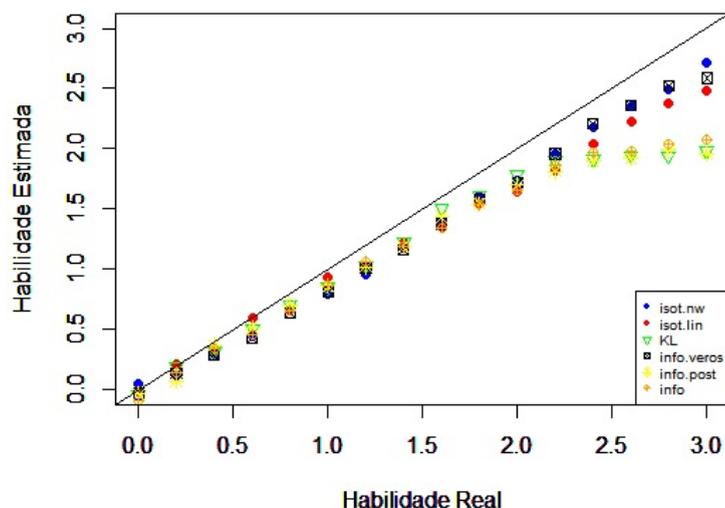
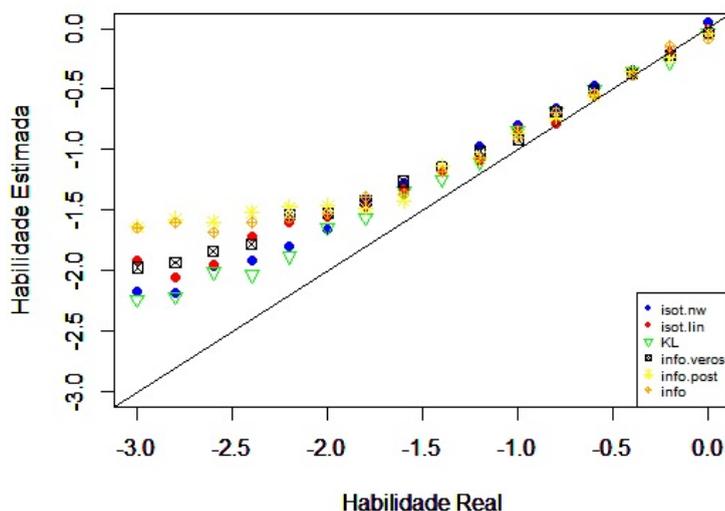


Figura 2.4: Habilidade simulada x Habilidade estimada - Valores de  $\theta$  negativo - Desvio Padrão a Posteriori = 0.2



se positivamente o critério de Kullback-Leibler, principalmente para  $\theta < -2$ . Os critérios propostos neste trabalho vêm logo depois do critério de Kullback-Leibler (verde) na proximidade da reta, juntamente com o critério do produto da verossimilhança com a informação de Fisher (preto).

Para os dois critérios propostos aqui, será apresentada uma tabela resumo, tal qual a Tabela 2.2, que traz o valor real de  $\theta$ , a média de  $\hat{\theta}$  estimado, o desvio padrão, o Erro Quadrático Médio (EQM), a média de itens utilizados e o desvio padrão no número de itens utilizando o critério de seleção do item pela Regressão Isotônica suavizada por Nadaraya-

Watson.

A medidas mostradas na Tabela 2.2, referentes a estimação de  $\theta$  são dadas por:

$$\widehat{E(\hat{\theta}_j)} = \frac{\sum_{l=1}^{100} \hat{\theta}_{lj}}{l},$$

$$\widehat{dp(\hat{\theta}_j)} = \frac{\sum_{l=1}^v [\hat{\theta}_{lj} - E(\hat{\theta}_j)]^2}{l},$$

$$\widehat{EQM(\hat{\theta}_j)} = [\theta_j - E(\hat{\theta}_j)]^2 + dp(\hat{\theta}_j)^2,$$

sendo  $\hat{\theta}_{lj}$  a habilidade estimada para a  $l$  – ésima simulação de  $\theta_j$ ,  $l=1,2,\dots,100$ .

Pode-se notar, além dos fatos observados nos gráficos anteriores, que o desvio padrão de  $\hat{\theta}$  é menor para os valores de  $\theta$  positivo, bem como o EQM, influenciado por essa medida.

Tabela 2.2: Tabela Resumo da Simulação da Isotônica Suavizada por Nadaraya-Watson  
- Desvio Padrão a Posteriori = 0.2

$\theta$	$\widehat{E(\hat{\theta})}$	$\widehat{dp(\hat{\theta})}$	$\widehat{EQM(\hat{\theta})}$	$\widehat{E(i)}$	$\widehat{dp(i)}$
-3	-2,2	0,72	1,21	6,5	1,53
-2,8	-2,2	0,49	0,62	6,6	1,52
-2,6	-2,0	0,52	0,67	6,5	1,28
-2,4	-1,9	0,51	0,50	6,3	1,11
-2,2	-1,8	0,45	0,37	6,5	1,35
-2	-1,7	0,43	0,31	6,4	1,51
-1,8	-1,4	0,48	0,37	6,4	1,07
-1,6	-1,3	0,42	0,28	6,5	1,40
-1,4	-1,2	0,42	0,23	6,7	1,53
-1,2	-1,0	0,41	0,22	6,9	1,50
-1	-0,8	0,40	0,21	7,2	1,57
-0,8	-0,7	0,43	0,20	7,2	1,55
-0,6	-0,5	0,35	0,14	7,4	1,43
-0,4	-0,3	0,37	0,14	7,4	1,90
-0,2	-0,2	0,33	0,11	7,6	1,66
0	0,0	0,34	0,12	7,8	1,83
0,2	0,1	0,41	0,17	8,2	2,05

0,4	0,3	0,31	0,11	8,3	2,28
0,6	0,4	0,38	0,17	8,6	2,24
0,8	0,7	0,36	0,15	9,1	2,20
1	0,8	0,28	0,12	8,9	2,00
1,2	1,0	0,36	0,19	9,3	2,17
1,4	1,2	0,29	0,13	10,3	2,37
1,6	1,4	0,30	0,14	10,7	2,17
1,8	1,6	0,33	0,14	11,0	2,52
2	1,7	0,29	0,15	11,0	2,10
2,2	2,0	0,28	0,14	11,7	2,18
2,4	2,2	0,25	0,11	12,5	2,45
2,6	2,4	0,26	0,13	12,8	2,16
2,8	2,5	0,21	0,14	13,3	2,27
3	2,7	0,27	0,15	14,3	2,50

A Tabela 2.3 trata da mesma informação que a Tabela 2.2, mas no uso da regressão isotônica com interpolação linear.

Também observa-se uma redução no desvio padrão de  $\hat{\theta}$  estimado para os valores positivos, bem como do EQM. Comparativamente ao critério suavizado por Nadaraya-Watson, o critério da isotônica com interpolação linear apresenta desvio padrão de  $\theta$  maior para os valores de  $\theta > 0$  e, conseqüentemente, um EQM maior também.

Tabela 2.3: Tabela Resumo da Simulação da Isotônica com Interpolação Linear - Desvio Padrão a Posteriori = 0.2

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-1,9	0,66	1,63	6,7	2,08
-2,8	-2,1	0,51	0,82	6,6	1,64
-2,6	-1,9	0,53	0,70	6,4	1,79
-2,4	-1,7	0,62	0,85	6,5	1,56
-2,2	-1,6	0,59	0,71	6,5	1,46
-2	-1,6	0,49	0,44	6,5	1,35
-1,8	-1,5	0,40	0,27	6,5	1,28

-1,6	-1,3	0,53	0,36	7	1,48
-1,4	-1,2	0,45	0,25	6,9	1,72
-1,2	-1,1	0,36	0,15	6,8	1,31
-1	-0,8	0,37	0,16	7,5	1,59
-0,8	-0,8	0,35	0,12	7,7	2,1
-0,6	-0,5	0,35	0,13	7,8	2,07
-0,4	-0,4	0,35	0,12	7,9	2,44
-0,2	-0,2	0,33	0,11	7,9	1,91
0	0,0	0,39	0,15	8	2,19
0,2	0,2	0,32	0,10	8,3	2,21
0,4	0,3	0,33	0,12	8,3	2,13
0,6	0,6	0,34	0,12	8,3	2,2
0,8	0,7	0,37	0,15	8,3	2,18
1	0,9	0,37	0,14	8,5	2,85
1,2	1,0	0,33	0,15	8,7	2,5
1,4	1,2	0,32	0,13	8,3	2,11
1,6	1,3	0,32	0,16	8,7	2,75
1,8	1,5	0,32	0,16	8,8	2,59
2	1,6	0,40	0,28	8,7	1,76
2,2	1,8	0,45	0,35	9,1	1,77
2,4	2,0	0,39	0,28	9,7	1,69
2,6	2,2	0,38	0,28	10,3	1,98
2,8	2,4	0,38	0,33	11,1	2,66
3	2,5	0,42	0,44	11,4	2,82

As tabelas resumo dos outros critérios podem ser visualizadas no Apêndice B. Todas as tabelas, exceto do critério de Kullback-Leibler, revelam um EQM maior para os valores de  $\theta > 0$ , assim como nos dois critérios aqui propostos. Entretanto, uma comparação do EQM para habilidades positivas do critério da regressão isotônica suavizada por Nadaraya-Watson com os outros critérios, mostra que o critério novo proposto apresenta resultados vantajosos, exceto quando comparado com o critério da verossimilhança ponderada pela informação de Fisher,

## Desvio Padrão a Posteriori = 0.1

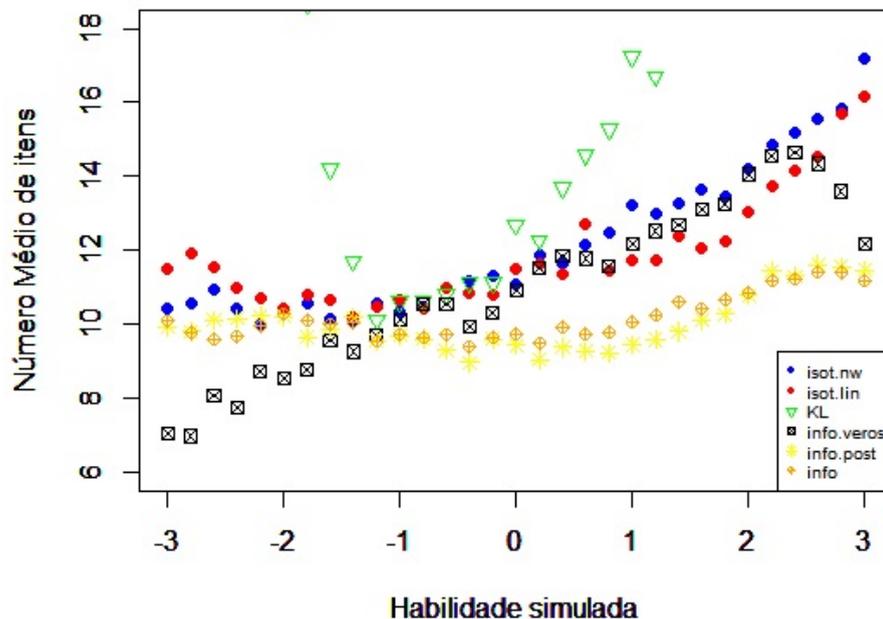
Em novo cenário, com desvio padrão a posteriori = 0.1, ou seja, com mais rigor no critério de convergência, foram aplicados mais itens, em média, do que o anterior, mantendo-se o limite máximo de 50 itens respondidos. Assim como no cenário anterior, não foi utilizado em momento algum os 50 itens máximos, sempre havendo convergência por causa do erro padrão definido como critério de parada.

Assim como no cenário anterior (Desvio Padrão a Posteriori = 0.2), a utilização do critério de Kullback-Leibler não produziu bons resultados, ficando com média de itens utilizados próximo aos outros critérios apenas no intervalo (-1;0), conforme Figura 2.5.

O comportamento de aumentar o número médio de itens utilizados à medida que se aumenta a habilidade continua para todos os critérios, mas menos acentuado no produto da posteriori com a informação de Fisher e no critério da máxima informação.

O critério do produto da verossimilhança pela informação de Fisher apresentou bons resultados para habilidades extremamente negativas, mas no intervalo (-1;2) apresentou resultados semelhantes aos dois critérios apresentados aqui.

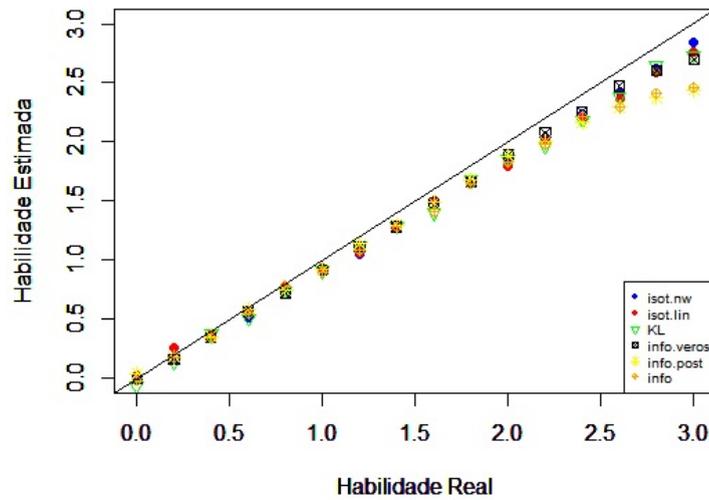
Figura 2.5: Quantidade média de itens utilizados em cada critério de escolha - Desvio Padrão a Posteriori = 0.1



Por fim, novamente o método da interpolação se mostra mais eficiente quando comparado apenas aos dois critérios novos para valores positivos de  $\theta$ .

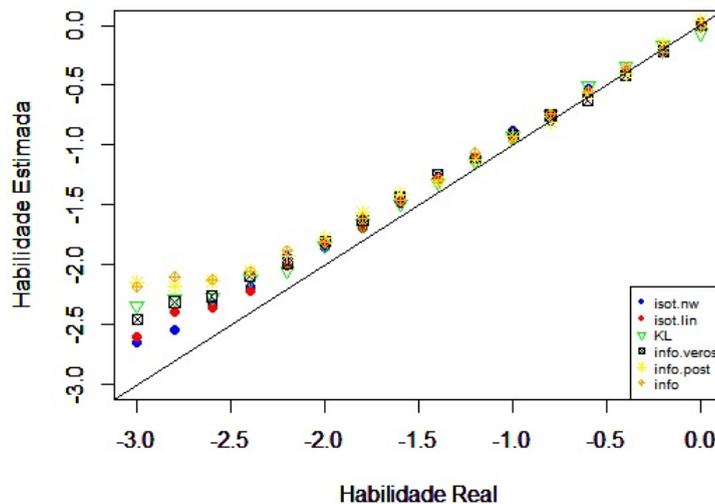
A análise do viés para valores de  $\theta$  positivo é semelhante ao cenário anterior. Todos os critérios apresentam acurácia boa para valores mais próximos do 0, mas na cauda os critérios que utilizam a máxima informação e o produto da posteriori com a informação de Fisher se distanciam da linha reta, mostrando ser menos acurados que os outros neste aspecto, conforme Figura 2.6.

Figura 2.6: Habilidade simulada x Habilidade estimada - Valores de  $\theta$  positivo - Desvio Padrão a Posteriori = 0.1



Para os valores de  $\theta$  negativo, repete-se o que se vê no valores positivos de  $\theta$ , com a pequena exceção de que o critério de Kullback-Leibler também se afasta da reta nos valores mais extremos.

Figura 2.7: Habilidade simulada x Habilidade estimada - Valores de  $\theta$  negativo - Desvio Padrão a Posteriori = 0.1



Pela Tabela 2.4 é possível notar que, assim como no cenário anterior, o EQM é menor para os valores positivos de  $\theta$  do que para os valores negativos e a Tabela 2.5 revela o mesmo para o critério da regressão isotônica com interpolação linear. A avaliação comparativa também se mantém a mesma: o EQM de  $\hat{\theta}$  é menor no critério de Nadaraya-Watson do que nos outros critérios, quando avaliado  $\theta > 0$  e excetuando o critério da verossimilhança ponderada pela informação.

Tabela 2.4: Tabela Resumo da Simulação da Isotônica Suavizada por Nadaraya-Watson  
- Desvio Padrão a Posteriori = 0.1

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,6	0,37	0,26	10,5	2,29
-2,8	-2,5	0,28	0,14	10,6	2,45
-2,6	-2,3	0,30	0,16	10,9	2,50
-2,4	-2,2	0,38	0,19	10,4	2,57
-2,2	-1,9	0,37	0,21	10,0	2,53
-2	-1,9	0,27	0,10	10,4	2,10
-1,8	-1,6	0,35	0,15	10,6	2,35
-1,6	-1,5	0,29	0,10	10,1	1,82
-1,4	-1,3	0,30	0,11	10,1	2,07
-1,2	-1,1	0,30	0,10	10,6	2,32
-1	-0,9	0,35	0,14	10,3	1,89
-0,8	-0,8	0,28	0,08	10,5	2,11
-0,6	-0,5	0,26	0,07	11,0	2,69
-0,4	-0,4	0,24	0,06	11,2	2,97
-0,2	-0,2	0,30	0,09	11,3	2,99
0	0,0	0,25	0,06	11,1	2,59
0,2	0,2	0,28	0,08	11,9	3,17
0,4	0,3	0,26	0,07	11,7	2,75
0,6	0,5	0,26	0,07	12,2	2,51
0,8	0,7	0,31	0,10	12,5	3,17
1	0,9	0,22	0,06	13,2	3,43
1,2	1,0	0,27	0,10	13,0	2,72
1,4	1,3	0,26	0,08	13,3	2,46

1,6	1,5	0,24	0,07	13,7	2,24
1,8	1,6	0,24	0,08	13,5	2,15
2	1,9	0,22	0,07	14,2	2,46
2,2	2,0	0,23	0,08	14,9	2,67
2,4	2,2	0,22	0,08	15,2	2,55
2,6	2,4	0,24	0,09	15,6	2,28
2,8	2,6	0,23	0,09	15,8	2,38
3	2,8	0,19	0,06	17,2	2,61

Tabela 2.5: Tabela Resumo da Simulação da Isotônica com Interpolação Linear - Desvio Padrão a Posteriori = 0.1

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,6	0,46	0,38	11,5	3,08
-2,8	-2,4	0,46	0,37	11,9	3,53
-2,6	-2,4	0,40	0,23	11,6	3,67
-2,4	-2,2	0,32	0,14	11,0	2,97
-2,2	-2,0	0,37	0,17	10,7	2,42
-2	-1,8	0,26	0,10	10,4	2,58
-1,8	-1,7	0,28	0,09	10,8	2,84
-1,6	-1,5	0,39	0,17	10,7	2,82
-1,4	-1,3	0,37	0,15	10,2	2,31
-1,2	-1,1	0,30	0,10	10,5	3,01
-1	-0,9	0,32	0,11	10,7	2,27
-0,8	-0,8	0,29	0,09	10,5	2,36
-0,6	-0,6	0,24	0,06	11,0	2,44
-0,4	-0,4	0,26	0,07	10,9	2,61
-0,2	-0,2	0,28	0,08	10,8	2,67
0	0,0	0,34	0,12	11,5	3,20
0,2	0,3	0,30	0,09	11,7	4,48
0,4	0,4	0,26	0,07	11,4	2,73
0,6	0,6	0,31	0,09	12,7	4,71

0,8	0,8	0,27	0,07	11,5	3,56
1	0,9	0,30	0,09	11,8	3,07
1,2	1,1	0,28	0,10	11,7	3,30
1,4	1,3	0,26	0,08	12,4	4,07
1,6	1,5	0,28	0,09	12,1	3,56
1,8	1,6	0,27	0,10	12,3	3,11
2	1,8	0,25	0,11	13,0	3,91
2,2	2,0	0,29	0,12	13,8	4,23
2,4	2,2	0,28	0,11	14,1	3,84
2,6	2,4	0,29	0,14	14,5	3,41
2,8	2,6	0,29	0,13	15,7	3,33
3	2,8	0,25	0,12	16,1	3,58

## Número de itens fixo

Os novos critérios aqui propostos apresentaram um vício pequeno quando comparado aos outros para os valores extremos de  $\theta$ . O motivo pode ser por que para esse níveis de habilidade eles utilizam mais itens para obter a convergência.

Para avaliar essa hipótese, outra simulação foi feita com cada um dos critérios, sem a utilização do desvio padrão a posteriori, e fixando o número de itens a serem respondidos em 30 itens.

A Tabela 2.6 apresenta o viés médio para cada nível de habilidade simulado para cada um dos critérios aqui comparado: Regressão Isotônica suavizada por Nadaraya-Watson (RINW), Regressão Isotônica com Interpolação Linear (RIIL), Máxima Informação (MI), Verossimilhança Ponderada pela Informação (VPI), Posteriori Ponderada pela Informação (PPI) e Kullback-Leibler (KL). O viés apresentado é dado pela diferença  $\theta - E(\hat{\theta})$

Destaca-se que a Regressão Isotônica suavizada por Nadaraya-Watson possui o menor viés médio dentre todos os critérios comparados (com aproximação de duas casas decimais) para os valores de  $\theta$  igual a -1; -0,6; 1,2; 1,8; 2,4; 2,8 e 3. Já a Regressão Isotônica com Interpolação Linear possui o menor viés médio para os valores de  $\theta$  igual a -3; -2,8; -2,6; -2,4; -0,4; -0,2; 0 e 1.

Os critérios aqui propostos apresentam maior acurácia para valores extremos e resultados semelhantes para valores intermediários quando o tamanho do teste é pré-fixado.

Tabela 2.6: Viés Médio em cada nível de habilidade - Teste de Tamanho Fixo em 30 Itens.

$\theta$	RINW	RIIL	MI	VPI	PPI	KL
<b>-3</b>	-0,10	-0,07	-0,20	-0,16	-0,22	-0,71
<b>-2,8</b>	-0,07	-0,05	-0,14	-0,09	-0,13	-0,51
<b>-2,6</b>	-0,06	-0,05	-0,08	-0,05	-0,07	-0,35
<b>-2,4</b>	-0,06	-0,05	-0,06	-0,05	-0,05	-0,24
<b>-2,2</b>	-0,07	-0,05	-0,04	-0,04	-0,04	-0,10
<b>-2</b>	-0,04	-0,05	-0,02	-0,06	-0,05	-0,01
<b>-1,8</b>	-0,04	-0,05	-0,05	-0,03	-0,03	0,03
<b>-1,6</b>	-0,02	-0,03	-0,01	-0,05	-0,03	0,03
<b>-1,4</b>	-0,04	-0,04	-0,03	-0,04	0,00	-0,02
<b>-1,2</b>	-0,03	-0,02	-0,03	0,00	-0,03	-0,04

<b>-1</b>	0,00	-0,02	-0,02	-0,04	-0,01	-0,03
<b>-0,8</b>	-0,02	-0,02	-0,01	0,00	0,00	-0,05
<b>-0,6</b>	-0,01	-0,03	-0,03	0,01	0,02	-0,03
<b>-0,4</b>	-0,02	0,00	0,00	0,00	-0,02	-0,04
<b>-0,2</b>	-0,01	0,00	0,02	0,01	-0,01	0,00
<b>0</b>	-0,02	0,00	0,00	0,02	0,01	0,01
<b>0,2</b>	0,00	0,01	0,02	-0,02	0,01	0,00
<b>0,4</b>	0,02	0,02	0,02	0,02	0,00	0,02
<b>0,6</b>	0,01	0,02	0,00	0,03	0,02	0,00
<b>0,8</b>	0,03	0,03	0,00	0,03	0,01	0,03
<b>1</b>	0,01	0,00	0,02	-0,01	0,02	0,07
<b>1,2</b>	0,03	0,04	0,04	0,03	0,03	0,03
<b>1,4</b>	0,03	0,02	0,05	0,03	0,01	0,03
<b>1,6</b>	0,05	0,04	0,04	0,04	0,03	0,09
<b>1,8</b>	0,02	0,05	0,03	0,06	0,04	0,10
<b>2</b>	0,06	0,07	0,06	0,05	0,05	0,14
<b>2,2</b>	0,07	0,08	0,05	0,03	0,06	0,14
<b>2,4</b>	0,06	0,07	0,06	0,06	0,07	0,18
<b>2,6</b>	0,07	0,06	0,07	0,03	0,08	0,27
<b>2,8</b>	0,06	0,09	0,11	0,10	0,12	0,48
<b>3</b>	0,07	0,10	0,17	0,16	0,17	0,60

A Tabela 2.7 apresenta o Erro Quadrático Médio em cada nível de habilidade, para cada critério proposto. Observa-se que os valores são bem próximos, exceto para o critério de Kullback-Leibler que apresenta um EQM destoante para  $\theta < -2,4$  e  $\theta > 1$ .

Tabela 2.7: EQM em cada nível de habilidade - Teste de Tamanho Fixo em 30 Itens.

$\theta$	RINW	RIL	MI	VPI	PPI	KL
<b>-3</b>	0,04	0,04	0,05	0,03	0,07	0,54
<b>-2,8</b>	0,03	0,04	0,05	0,03	0,05	0,30
<b>-2,6</b>	0,02	0,03	0,03	0,02	0,03	0,15
<b>-2,4</b>	0,03	0,03	0,03	0,03	0,02	0,10

<b>-2,2</b>	0,04	0,03	0,03	0,02	0,03	0,05
<b>-2</b>	0,02	0,03	0,02	0,04	0,03	0,04
<b>-1,8</b>	0,02	0,03	0,03	0,02	0,03	0,05
<b>-1,6</b>	0,03	0,03	0,02	0,03	0,02	0,04
<b>-1,4</b>	0,03	0,02	0,03	0,04	0,02	0,04
<b>-1,2</b>	0,03	0,03	0,02	0,02	0,02	0,03
<b>-1</b>	0,02	0,02	0,02	0,02	0,01	0,04
<b>-0,8</b>	0,02	0,02	0,02	0,02	0,02	0,02
<b>-0,6</b>	0,02	0,02	0,02	0,02	0,02	0,03
<b>-0,4</b>	0,02	0,02	0,02	0,02	0,02	0,03
<b>-0,2</b>	0,02	0,02	0,02	0,02	0,02	0,04
<b>0</b>	0,03	0,02	0,02	0,02	0,02	0,04
<b>0,2</b>	0,03	0,02	0,02	0,03	0,02	0,04
<b>0,4</b>	0,02	0,02	0,02	0,03	0,02	0,05
<b>0,6</b>	0,02	0,02	0,02	0,02	0,02	0,05
<b>0,8</b>	0,02	0,03	0,02	0,02	0,02	0,05
<b>1</b>	0,02	0,03	0,02	0,02	0,02	0,07
<b>1,2</b>	0,02	0,03	0,02	0,02	0,02	0,06
<b>1,4</b>	0,03	0,03	0,02	0,02	0,02	0,05
<b>1,6</b>	0,02	0,03	0,02	0,02	0,03	0,06
<b>1,8</b>	0,02	0,03	0,03	0,02	0,02	0,09
<b>2</b>	0,03	0,03	0,02	0,03	0,02	0,14
<b>2,2</b>	0,03	0,03	0,03	0,02	0,03	0,13
<b>2,4</b>	0,03	0,04	0,02	0,02	0,02	0,13
<b>2,6</b>	0,03	0,03	0,02	0,02	0,02	0,13
<b>2,8</b>	0,03	0,04	0,03	0,02	0,03	0,31
<b>3</b>	0,03	0,04	0,03	0,03	0,04	0,39

## Tempo de Simulação

Para registrar o tempo de simulação, foi simulado um teste de tamanho fixo de 30 itens para sete respondentes. Com valores de  $\theta = (-3, -2, -1, 0, 1, 2, 3)$ .

Os critérios mais demorados foram aqueles que utilizam a ponderação da informação de Fisher. O critério do produto da verossimilhança com a informação demorou 2 minutos e 58 segundos, aproximadamente. Já o critério do produto da posteriori com a informação demorou por volta de 2 minutos e 50 segundos.

Logo em seguida, ficou o critério que utiliza a informação de Kullback-Leibler, com tempo para simulação aproximado de 45 segundos.

Os critérios que utilizam a regressão isotônica aqui apresentados e o da máxima informação foram os mais rápidos. A regressão isotônica com interpolação linear durou 10 segundos, a máxima informação durou 7 segundos e a regressão isotônica suavizada por Nadaraya-Watson durou 3 segundos.

## 2.2 Banco de Itens Real

Além das simulações, para testar os novos critérios propostos, foi feita uma aplicação num banco de itens real. As respostas serão simuladas tal qual na seção anterior, mas a diferença é que ao invés daqueles 100 bancos simulados, aqui serão utilizados 383 itens do banco da prova de Proficiência em Inglês da Universidade de Brasília (UnB).

A Figura 2.8 mostra a distribuição dos parâmetros  $a$  e  $b$  dos itens utilizados. Nota-se que não há itens com  $b > 2.5$ , o que prejudicará a estimativa para valores de  $\theta$  dessa magnitude.

Aqui serão simuladas aplicações para 50 respondentes em cada nível de habilidade, variando de -2 a 2, com incrementos de 0.2 entre eles. O intervalo menor se deve ao fato de termos poucos  $b$  maiores que 2, como mostrado na Figura 2.8.

O critério de parada utilizado foi o Desvio Padrão a Posteriori = 0.1 e/ou o número máximo de 20 itens respondidos, ou seja, todo teste.

A Figura 2.9 mostra a quantidade média de itens utilizados em cada critério. O critério do produto da posteriori com a informação de Fisher foi o melhor para todo  $\theta$ .

Os novos critérios aqui propostos apresentaram resultados semelhantes ao produto da verossimilhança com a informação e ao critério da máxima informação.

A análise do viés, na Figura 2.10, mostra que os critérios possuem boa estimativa até  $\theta = 1.2$  e resultados muito semelhantes entre eles.

Para valores negativos de  $\theta$  vale o mesmo observado para valores positivos, ou seja, os critérios apresentam boa estimativa para  $\theta > -1.2$ .

Figura 2.8: Histograma dos Parâmetros de Discriminação e de Dificuldade - Proficiência em Inglês da UnB

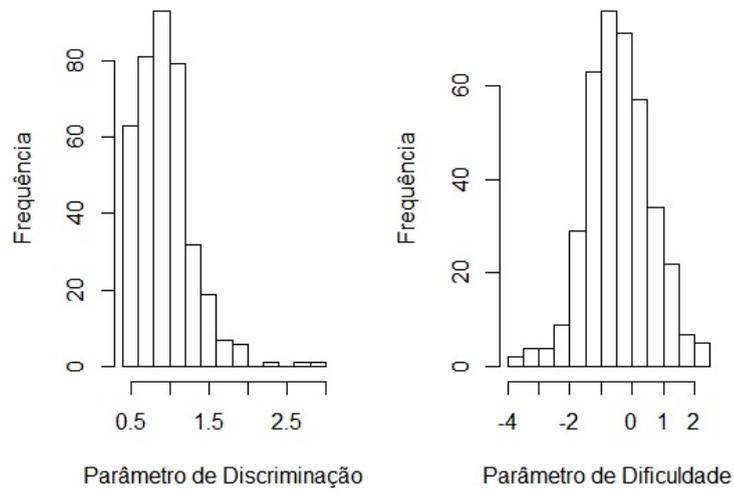


Figura 2.9: Quantidade média de itens utilizados em cada critério de escolha - Proficiência em Inglês da UnB

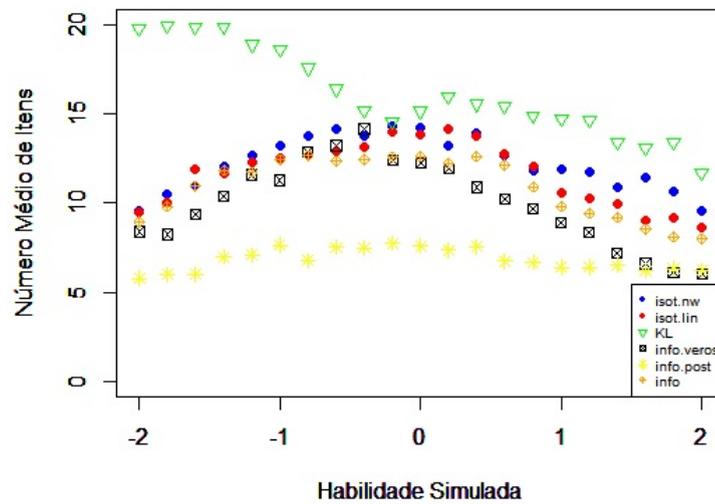


Figura 2.10: Habilidade simulada x Habilidade estimada - Valores de  $\theta$  positivo - Proficiência em Inglês da UnB

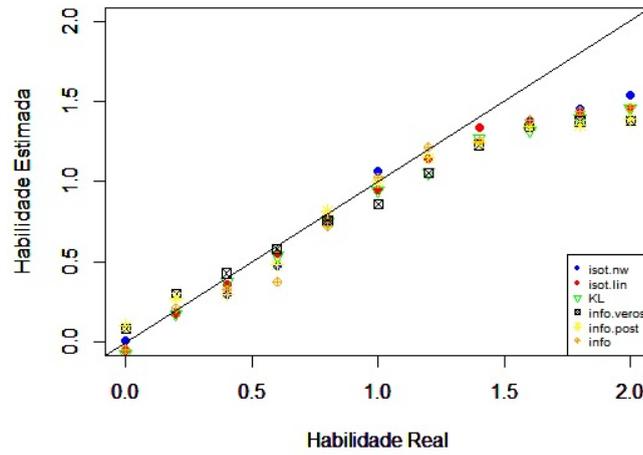
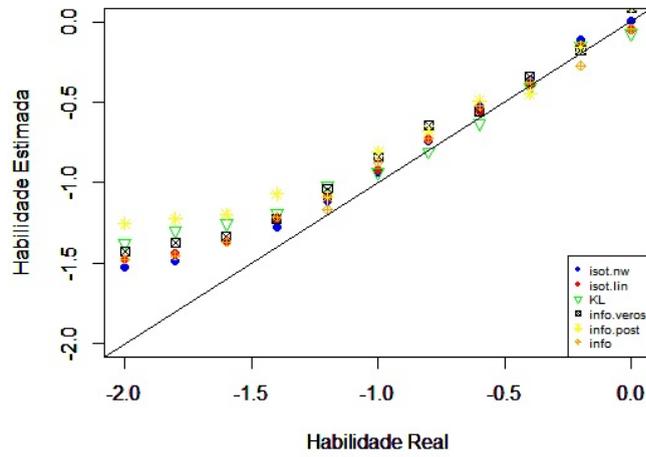


Figura 2.11: Habilidade simulada x Habilidade estimada - Valores de  $\theta$  positivo - Proficiência em Inglês da UnB



## Capítulo 3

# Conclusão e Propostas

Os estudos de simulação mostraram que os dois critérios propostos apesar de apresentarem bons resultados não se qualificam, ainda, para substituir os métodos já costumeiramente utilizados. Nos cenários com o desvio-padrão a posteriori como critério de parada ambos os critérios mostraram necessitar de mais itens para obter a convergência e, em contrapartida, maior acurácia principalmente para os valores extremos.

Acreditava-se que a maior acurácia pudesse ser consequência da maior quantidade de itens respondidos, mas o cenário de teste de tamanho fixo manteve a mesma perspectiva de maior acurácia nas extremidades com os dois critérios propostos com base na regressão isotônica

A aplicação na prova de Proficiência de Inglês da UnB revelou que os critérios novos necessitam de maiores ajustes para banco de itens menores.

Para aplicação mais rápida do teste e com um banco de itens relativamente grande, os critérios propostos podem surgir como alternativa, já que os resultados são semelhantes e a velocidade de processamento é mais rápida, principalmente no uso da regressão isotônica suavizada por Nadaraya-Watson.

Como proposta para estudos futuros, existe a possibilidade de testar algumas mudanças nos critérios apresentados. O critério que utiliza a regressão isotônica suavizada por Nadaraya-Watson, por exemplo, possui vários fatores que, se alterados, podem apresentar resultados bem distintos dos que aqui foram apresentados.

O primeiro fator a ser analisado são os valores do parâmetro  $b$  incluídos nas extremidades da regressão isotônica, conforme Figura 2.1. Estudos empíricos dão evidências de que valores menores do que o incremento ou decréscimo de uma unidade do máximo e

mínimo de  $b$ 's já respondidos, podem melhorar o critério de escolha do item. Um estudo de otimização permitirá identificar qual incremento será ótimo para o critério.

Outro fator bastante influente nos critérios propostos é o valor estabelecido de 0.6 a ser buscado na variável resposta da regressão isotônica. Mudança nesse valor, inclusive tornando-o variável conforme o valor do  $\theta$  atualmente estimado, pode melhorar a convergência. Além disso, na suavização por Nadaraya-Watson foi escolhida uma distribuição normal com parâmetro de suavização  $h = 1$ . Alterações na função kernel e no parâmetro  $h$  também podem alterar a convergência do critério proposto.

Além da tentativa de melhorar os critérios aqui apresentados, há uma necessidade também de um estudo mais específico de impacto na exposição do banco de itens de cada critério comparado.

Por fim, há a possibilidade de estudar o uso do  $\theta$  encontrado na regressão isotônica como estimativa do verdadeiro  $\theta$ , não somente para a seleção do próximo item. Seria uma proposta de seleção do item e estimativa do  $\theta$  utilizando a mesma regressão isotônica.

# Referências Bibliográficas

- Andrade, D. F., Tavares, H. R., e Valle, R. C. (2000). Teoria da resposta ao item: conceitos e aplicações. *ABE, Sao Paulo*.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4):561–573.
- Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., Silverman, E., et al. (1955). An empirical distribution function for sampling with incomplete information. *The annals of mathematical statistics*, 26(4):641–647.
- Barlow, R. E., Bartholomew, D. J., Bremner, J., e Brunk, H. D. (1972). *Statistical inference under order restrictions: The theory and application of isotonic regression*. Wiley New York.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, pages 395–479.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1):29–51.
- Bock, R. D. e Mislevy, R. J. (1982). Adaptive eap estimation of ability in a microcomputer environment. *Applied psychological measurement*, 6(4):431–444.
- Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software*, 71(5):1–39.
- Chang, H.-H. e van der Linden, W. J. (2003). Optimal stratification of item pools in  $\alpha$ -stratified computerized adaptive testing. *Applied Psychological Measurement*, 27(4):262–274.

- Chang, H.-H. e Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3):213–229.
- Chang, H.-H. e Ying, Z. (1999).  $\alpha$ -stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23(3):211–222.
- Chang, H.-H. e Ying, Z. (2008). To weight or not to weight? balancing influence of initial items in adaptive testing. *Psychometrika*, 73(3):441–450.
- Chang, H.-H. e Ying, Z. (2009). Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *The Annals of Statistics*, pages 1466–1488.
- Choe, E. M., Kern, J. L., e Chang, H.-H. (2017). Optimizing the use of response times for item selection in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 43(2):135–158.
- Costa, D. R. (2009). Métodos estatísticos em testes adaptativos informatizados. *Master's thesis, UFRJ-Universidade Federal do Rio de Janeiro*, 9.
- Fan, Z., Wang, C., Chang, H.-H., e Douglas, J. (2012). Utilizing response time distributions for item selection in cat. *Journal of Educational and Behavioral Statistics*, 37(5):655–670.
- Gray, R. (2001). Bio 248 cd advanced statistical computing.
- Hambleton, R. K., Swaminathan, H., e Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Härdle, W. (1990). *Applied nonparametric regression*. Number 19. Cambridge university press.
- Kaplan, M., de la Torre, J., e Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied psychological measurement*, 39(3):167–188.
- Klaus, B. e Strimmer, K. (2015). *fdrtool: Estimation of (Local) False Discovery Rates and Higher Criticism*. R package version 1.2.15.
- Labarrere, J. G., DA-SILVA, C. Q., e COSTA, D. R. (2011). Testes adaptativos computadorizados. *Revista Brasileira de Biometria, São Paulo*, 29(2):229–261.

- Lee, Y.-S. (2007). A comparison of methods for nonparametric estimation of item characteristic curves for binary items. *Applied Psychological Measurement*, 31(2):121–134.
- Lord, F. (1952). A theory of test scores. *Psychometric monographs*.
- Lord, F. M. (1986). Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 23(2):157–162.
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2):149–174.
- Mislevy, R. J. (1986a). Bayes modal estimation in item response models. *Psychometrika*, 51(2):177–195.
- Mislevy, R. J. (1986b). Recent developments in the factor analysis of categorical variables. *Journal of Educational and Behavioral Statistics*, 11(1):3–31.
- Moreira Junior, F. J. (2010). Aplicações da teoria da resposta ao item (tri) no brasil. *Rev. Bras. Biom*, 28(4):137–170.
- Muraki, E. (1997). A generalized partial credit model. In: *Handbook of modern item response theory*, pages 153–164. Springer.
- Pereira, E. A. (2014). Algumas propostas para imputação de dados faltantes em teoria de resposta ao item.
- R Core Team. *R: A Language and Environment for Statistical Computing*.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.
- Sands, W. A. e Waters, B. K. (1997). Introduction to asvab and cat. *Computerized adaptive testing: From inquiry to operation*, pages 3–9.
- Schlemper, B. N. (2014). Testes de adequabilidade de ajuste em teoria de resposta ao item. Master's thesis, Universidade de Brasília.
- van der Linden, W. J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63(2):201–216.

- Van der Linden, W. J. e Pashley, P. J. (2009). Item selection and ability estimation in adaptive testing. In: *Elements of adaptive testing*, pages 3–30. Springer.
- Veerkamp, W. J. e Berger, M. P. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, 22(2):203–226.
- Wainer, H. (2000). *Computerized adaptive testing*. Wiley Online Library.
- Wang, T., Hanson, B. A., e Lau, C.-M. A. (1999). Reducing bias in cat trait estimation: A comparison of approaches. *Applied Psychological Measurement*, 23(3):263–278.
- Wang, T. e Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(2):109–135.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3):427–450.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied psychological measurement*, 6(4):473–492.
- Weiss, D. J. e McBride, J. R. (1984). Bias and information of bayesian adaptive testing. *Applied Psychological Measurement*, 8(3):273–285.
- Wirth, R. e Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychological methods*, 12(1):58.

# Apêndice A

## Quadratura Gaussiana

Algumas integrais sem solução analítica, como aquelas utilizadas para estimar  $\theta$ , via EAP, e o desvio a posteriori associado ao EAP, exigem métodos de integração numérica. Para estes casos, na TRI, usualmente recorre-se ao método da Quadratura Gaussiana, que nada mais é do que uma aproximação da integral pela soma de áreas de um número finito de retângulos.

Gray (2001) apresenta a forma básica da regra da Quadratura Gaussiana como:

$$\int_a^b W(x)f(x)dx = \sum_{j=1}^n \omega_j f(x_j) \quad (\text{A.1})$$

de tal forma que os pesos  $\omega_j$  e as abscissas  $x_j$  são definidos de forma que se obtenha uma aproximação precisa. Com  $\omega_j$  e  $x_j$  bem definidos, é possível obter a fórmula exata se  $f(x)$  for um polinômio de grau  $2n - 1$ .

Para determinar os pesos e as abscissas usa-se um sistema de polinômios ortogonais. Sendo  $\{u(x) : \int_a^b W(x)u(x)^2 dx < \infty\}$  um espaço funcional, então um polinômio base  $\{p_j(x), j = 0, 1, 2, \dots\}$  pode ser dado para este espaço funcional, de tal forma que  $p_j(x)$  é um polinômio de grau  $j$  e  $p_j(x)$  é ortogonal, no sentido de que  $\int_a^b W(x)p_j(x)p_k(x)dx = 0$  para  $j \neq k$ . Há fórmulas recorrentes simples para  $p_j(x)$ , como  $p_0(x) \propto 1$ . Será demonstrado que, para todo  $p_j(x)$ , haverá  $j$  raízes reais distintas.

Dado um conjunto de possíveis abscissas  $x_1, \dots, x_n$  para (A.1), os pesos  $w_j, j = 1, \dots, n$ , podem ser definidos para tornar (A.1) exata para as funções  $p_0(x), \dots, p_{n-1}(x)$ . A princípio

isso pode ser feito pela solução do sistema de equações

$$\int_a^b W(x)p_i(x)dx = \sum_{j=1}^n \omega_j p_i(x_j), \quad i = 0, \dots, n-1,$$

O sistema tende a ser mal-condicionado e, na prática, há métodos melhores que a solução direta dele. Os valores da abscissa podem variar tornando (A.1) exato para  $p_j(x), j = n, \dots, 2n-1$  também. As raízes dos polinômios sempre estarão no interior do intervalo, então o integrando não necessita ter valor nos limites de integração.

# Apêndice B

## Tabelas Resumo da Simulação

Tabela B.1: Tabela Resumo da Simulação da Máxima Informação - Desvio Padrão a Posteriori = 0.2

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-1,6	0,5	2,08	5,8	0,86
-2,8	-1,6	0,51	1,68	5,9	0,98
-2,6	-1,7	0,44	1,03	6,0	1,04
-2,4	-1,6	0,54	0,94	6,1	1,27
-2,2	-1,6	0,45	0,61	6,0	0,99
-2	-1,5	0,45	0,42	6,1	1,01
-1,8	-1,4	0,53	0,45	6,0	0,98
-1,6	-1,4	0,49	0,3	6,4	1,39
-1,4	-1,2	0,44	0,24	6,3	1,11
-1,2	-1,1	0,4	0,18	6,3	0,78
-1	-0,9	0,39	0,16	6,5	0,93
-0,8	-0,7	0,37	0,15	6,5	0,98
-0,6	-0,5	0,37	0,14	6,7	0,98
-0,4	-0,4	0,37	0,14	6,5	0,79
-0,2	-0,1	0,33	0,11	6,7	0,87
0	-0,1	0,35	0,13	6,6	0,65
0,2	0,1	0,37	0,14	6,9	1,05
0,4	0,3	0,28	0,08	6,8	0,81

0,6	0,5	0,34	0,14	7,0	0,85
0,8	0,7	0,34	0,13	7,1	0,87
1	0,9	0,36	0,15	7,2	0,82
1,2	1,1	0,34	0,13	7,2	0,85
1,4	1,2	0,32	0,15	7,5	0,87
1,6	1,3	0,32	0,16	7,5	0,74
1,8	1,5	0,36	0,2	7,7	0,64
2	1,7	0,34	0,23	7,8	0,60
2,2	1,9	0,31	0,2	7,9	0,35
2,4	2	0,26	0,25	7,9	0,37
2,6	2	0,24	0,44	8,1	0,35
2,8	2	0,19	0,62	8,0	0,14
3	2,1	0,08	0,86	8,0	0,22

Tabela B.2: Tabela Resumo da Simulação da Máxima Informação Ponderada pela Posteriori - Desvio Padrão a Posteriori = 0.2

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-1,6	0,51	2,12	5,7	1,22
-2,8	-1,6	0,50	1,76	5,7	1,12
-2,6	-1,6	0,53	1,28	5,6	1,13
-2,4	-1,5	0,51	1,03	5,9	1,72
-2,2	-1,5	0,56	0,83	6,0	1,42
-2	-1,5	0,49	0,53	6,0	1,29
-1,8	-1,5	0,43	0,29	5,9	1,24
-1,6	-1,4	0,36	0,16	6,0	1,61
-1,4	-1,1	0,44	0,26	6,1	1,38
-1,2	-1,1	0,34	0,13	6,1	1,45
-1	-0,9	0,44	0,21	6,1	1,47
-0,8	-0,7	0,41	0,17	5,8	0,99
-0,6	-0,6	0,37	0,14	6,0	1,14
-0,4	-0,4	0,35	0,12	6,3	1,21
-0,2	-0,2	0,34	0,12	6,1	0,94
0	0,0	0,36	0,13	6,2	1,01
0,2	0,1	0,31	0,12	6,0	0,79
0,4	0,4	0,35	0,13	6,4	1,10
0,6	0,5	0,35	0,13	6,6	1,00
0,8	0,7	0,39	0,16	6,5	1,05
1	0,9	0,33	0,13	6,7	0,96
1,2	1,0	0,33	0,14	6,9	0,94
1,4	1,2	0,33	0,15	7,0	0,97
1,6	1,4	0,35	0,15	7,3	0,81
1,8	1,5	0,36	0,20	7,4	0,72
2	1,7	0,38	0,23	7,7	0,74
2,2	1,8	0,30	0,24	7,6	0,56
2,4	1,9	0,25	0,32	7,7	0,58
2,6	1,9	0,23	0,51	7,7	0,51

2,8	2,0	0,14	0,68	7,7	0,48
3	2,0	0,17	1,10	7,7	0,53

Tabela B.3: Tabela Resumo da Simulação da Máxima Informação Ponderada pela Verossimilhança - Desvio Padrão a Posteriori = 0.2

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,0	0,56	1,37	4,6	2,42
-2,8	-1,9	0,56	1,05	4,5	1,96
-2,6	-1,8	0,53	0,86	4,9	2,89
-2,4	-1,8	0,51	0,64	5,4	3,33
-2,2	-1,5	0,64	0,84	5,3	1,95
-2	-1,5	0,55	0,52	5,7	3,26
-1,8	-1,4	0,47	0,36	5,4	1,76
-1,6	-1,3	0,37	0,25	5,8	2,04
-1,4	-1,1	0,45	0,27	6,1	2,16
-1,2	-1,0	0,44	0,23	5,9	1,58
-1	-0,9	0,38	0,15	6,4	1,76
-0,8	-0,7	0,39	0,16	6,3	1,57
-0,6	-0,5	0,32	0,11	7,0	2,40
-0,4	-0,4	0,32	0,10	7,4	2,45
-0,2	-0,2	0,39	0,15	7,8	2,82
0	0,0	0,33	0,11	7,8	2,70
0,2	0,1	0,30	0,10	8,4	3,57
0,4	0,3	0,33	0,12	8,4	2,77
0,6	0,4	0,31	0,13	8,5	2,94
0,8	0,6	0,33	0,13	8,9	2,95
1	0,8	0,31	0,13	8,7	2,17
1,2	1,0	0,28	0,11	9,5	2,84
1,4	1,2	0,32	0,16	9,8	2,24
1,6	1,4	0,28	0,13	10,5	2,85
1,8	1,6	0,29	0,13	10,5	2,85
2	1,7	0,28	0,16	11,0	3,00
2,2	2,0	0,22	0,11	11,6	3,23
2,4	2,2	0,28	0,12	12,2	2,64
2,6	2,4	0,27	0,13	11,8	2,40

2,8	2,5	0,20	0,12	11,2	1,75
3	2,6	0,21	0,21	10,5	1,43

Tabela B.4: Tabela Resumo da Simulação de Kullback-Leibler - Desvio Padrão a Posteriori = 0.2

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,2	0,41	0,75	25,5	21,83
-2,8	-2,2	0,34	0,47	21,9	21,21
-2,6	-2,0	0,49	0,59	17,8	19,58
-2,4	-2,0	0,38	0,28	19,4	20,50
-2,2	-1,9	0,45	0,30	13,3	16,64
-2	-1,6	0,38	0,27	8,0	9,68
-1,8	-1,6	0,41	0,23	7,7	8,18
-1,6	-1,3	0,34	0,19	7,0	6,25
-1,4	-1,2	0,38	0,17	6,1	1,04
-1,2	-1,1	0,29	0,09	6,5	1,43
-1	-0,8	0,33	0,14	6,7	1,41
-0,8	-0,7	0,35	0,13	7,2	1,80
-0,6	-0,5	0,35	0,13	7,3	1,86
-0,4	-0,4	0,35	0,12	7,9	2,23
-0,2	-0,3	0,35	0,13	7,6	2,14
0	0,0	0,32	0,11	8,5	2,58
0,2	0,2	0,28	0,08	9,5	2,90
0,4	0,3	0,31	0,10	9,6	2,94
0,6	0,5	0,34	0,13	10,4	3,21
0,8	0,7	0,37	0,15	11,4	3,22
1	0,9	0,29	0,10	12,1	2,82
1,2	1,0	0,34	0,14	14,0	3,17
1,4	1,2	0,40	0,19	15,2	3,13
1,6	1,5	0,45	0,21	17,1	2,96
1,8	1,6	0,48	0,26	17,5	2,84
2	1,8	0,38	0,19	18,6	1,82
2,2	1,9	0,29	0,17	18,9	1,65
2,4	1,9	0,27	0,31	18,9	1,77
2,6	1,9	0,25	0,50	19,1	1,51

2,8	1,9	0,24	0,80	19,1	1,28
3	2,0	0,15	1,06	19,3	1,10

Tabela B.5: Tabela Resumo da Simulação da Máxima Informação - Desvio Padrão a Posteriori = 0.1

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,2	0,41	0,84	10	2,65
-2,8	-2,1	0,48	0,72	10	2,87
-2,6	-2,1	0,52	0,50	10	2,39
-2,4	-2,1	0,50	0,37	10	2,28
-2,2	-1,9	0,43	0,28	10	2,20
-2	-1,8	0,38	0,18	10	2,14
-1,8	-1,6	0,39	0,19	10	2,01
-1,6	-1,5	0,29	0,11	10	2,30
-1,4	-1,3	0,35	0,13	10	2,06
-1,2	-1,1	0,27	0,09	10	1,87
-1	-1,0	0,32	0,10	10	1,77
-0,8	-0,7	0,26	0,07	10	1,42
-0,6	-0,5	0,27	0,08	10	1,52
-0,4	-0,4	0,23	0,06	9	1,33
-0,2	-0,2	0,27	0,07	10	1,42
0	0,0	0,24	0,06	10	1,57
0,2	0,2	0,31	0,10	9	1,03
0,4	0,3	0,26	0,07	10	1,58
0,6	0,5	0,30	0,09	10	1,22
0,8	0,8	0,24	0,06	10	1,08
1	0,9	0,30	0,10	10	1,37
1,2	1,1	0,29	0,10	10	1,57
1,4	1,3	0,29	0,09	11	1,63
1,6	1,4	0,28	0,11	10	1,21
1,8	1,6	0,27	0,09	11	1,33
2	1,8	0,29	0,11	11	1,34
2,2	2,0	0,35	0,17	11	1,60
2,4	2,2	0,27	0,10	11	1,34
2,6	2,3	0,22	0,14	11	1,39

2,8	2,4	0,14	0,17	11	1,25
3	2,5	0,08	0,29	11	0,81

Tabela B.6: Tabela Resumo da Simulação da Máxima Informação Ponderada pela Posteriori - Desvio Padrão a Posteriori = 0.1

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,1	0,51	0,99	9,9	2,84
-2,8	-2,2	0,41	0,55	9,8	3,04
-2,6	-2,1	0,37	0,36	10,1	2,71
-2,4	-2,0	0,43	0,30	10,2	2,65
-2,2	-1,9	0,39	0,23	10,2	2,93
-2	-1,8	0,39	0,20	10,2	2,67
-1,8	-1,6	0,41	0,23	9,7	2,16
-1,6	-1,4	0,34	0,15	9,9	2,36
-1,4	-1,3	0,31	0,11	10,2	2,67
-1,2	-1,1	0,28	0,09	9,6	2,42
-1	-0,9	0,33	0,11	9,7	2,55
-0,8	-0,8	0,26	0,07	9,6	2,48
-0,6	-0,6	0,32	0,10	9,3	2,10
-0,4	-0,4	0,25	0,06	9,0	1,31
-0,2	-0,2	0,24	0,06	9,6	1,82
0	0,0	0,33	0,11	9,5	1,82
0,2	0,2	0,29	0,08	9,1	1,35
0,4	0,4	0,25	0,07	9,4	1,68
0,6	0,6	0,26	0,07	9,3	1,39
0,8	0,7	0,24	0,06	9,2	1,02
1	0,9	0,26	0,07	9,5	1,33
1,2	1,1	0,25	0,06	9,6	1,17
1,4	1,3	0,28	0,09	9,8	1,22
1,6	1,5	0,29	0,10	10,1	1,73
1,8	1,7	0,30	0,10	10,3	1,13
2	1,9	0,30	0,11	10,8	1,56
2,2	2,0	0,28	0,12	11,5	1,95
2,4	2,2	0,27	0,13	11,3	1,37
2,6	2,3	0,23	0,14	11,6	1,59

2,8	2,4	0,21	0,23	11,6	1,54
3	2,4	0,13	0,33	11,5	0,85

Tabela B.7: Tabela Resumo da Simulação da Máxima Informação Ponderada pela Verossimilhança - Desvio Padrão a Posteriori = 0.1

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,5	0,20	0,34	7,1	3,59
-2,8	-2,3	0,45	0,45	7,0	2,34
-2,6	-2,3	0,38	0,25	8,1	4,23
-2,4	-2,1	0,43	0,28	7,8	2,46
-2,2	-2,0	0,34	0,16	8,7	3,25
-2	-1,8	0,41	0,21	8,6	2,65
-1,8	-1,6	0,40	0,19	8,8	1,99
-1,6	-1,4	0,38	0,17	9,6	2,78
-1,4	-1,2	0,35	0,15	9,3	2,19
-1,2	-1,1	0,31	0,11	9,7	2,91
-1	-0,9	0,29	0,09	10,1	3,23
-0,8	-0,7	0,27	0,08	10,5	3,01
-0,6	-0,6	0,28	0,08	10,6	2,73
-0,4	-0,4	0,28	0,08	9,9	2,96
-0,2	-0,2	0,27	0,07	10,3	2,55
0	0,0	0,29	0,08	10,9	3,17
0,2	0,2	0,23	0,05	11,5	3,38
0,4	0,3	0,28	0,08	11,9	3,50
0,6	0,6	0,23	0,05	11,8	3,26
0,8	0,7	0,27	0,08	11,6	2,25
1	0,9	0,23	0,06	12,2	2,86
1,2	1,1	0,21	0,05	12,5	3,54
1,4	1,3	0,22	0,06	12,7	3,03
1,6	1,5	0,25	0,08	13,1	3,17
1,8	1,7	0,26	0,09	13,3	2,93
2	1,9	0,20	0,05	14,1	3,08
2,2	2,1	0,24	0,07	14,5	2,88
2,4	2,3	0,22	0,07	14,7	3,49
2,6	2,5	0,18	0,05	14,4	2,41

2,8	2,6	0,13	0,06	13,6	2,29
3	2,7	0,12	0,10	12,2	1,93

Tabela B.8: Tabela Resumo da Simulação de Kullback-Leibler - Desvio Padrão a Posteriori = 0.1

$\theta$	$\widehat{E}(\hat{\theta})$	$\widehat{dp}(\hat{\theta})$	$\widehat{EQM}(\hat{\theta})$	$\widehat{E}(i)$	$\widehat{dp}(i)$
-3	-2,3	0,22	0,48	48,8	6,25
-2,8	-2,3	0,24	0,33	48,1	8,23
-2,6	-2,3	0,29	0,19	46,9	10,51
-2,4	-2,1	0,31	0,18	41,3	16,05
-2,2	-2,0	0,31	0,12	36,6	18,81
-2	-1,8	0,34	0,14	28,0	19,50
-1,8	-1,7	0,31	0,12	18,7	16,21
-1,6	-1,5	0,29	0,10	14,2	11,36
-1,4	-1,3	0,30	0,10	11,7	8,08
-1,2	-1,1	0,26	0,07	10,1	1,45
-1	-0,9	0,27	0,08	10,6	1,79
-0,8	-0,7	0,26	0,07	10,6	1,74
-0,6	-0,5	0,25	0,07	10,8	2,13
-0,4	-0,3	0,24	0,06	11,1	2,26
-0,2	-0,2	0,27	0,07	11,1	2,47
0	-0,1	0,28	0,08	12,7	3,78
0,2	0,1	0,30	0,09	12,3	3,70
0,4	0,4	0,29	0,09	13,7	5,33
0,6	0,5	0,29	0,09	14,6	6,08
0,8	0,7	0,29	0,08	15,3	6,10
1	0,9	0,28	0,09	17,2	7,77
1,2	1,1	0,26	0,07	16,7	7,66
1,4	1,3	0,33	0,12	20,0	8,54
1,6	1,4	0,28	0,12	20,2	8,08
1,8	1,7	0,29	0,10	26,2	9,66
2	1,9	0,31	0,12	28,1	8,87
2,2	2,0	0,36	0,18	29,9	8,33
2,4	2,2	0,29	0,13	34,5	8,26
2,6	2,4	0,34	0,17	39,2	7,94

2,8	2,6	0,39	0,17	43,7	7,22
3	2,7	0,35	0,20	45,5	6,67