



UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

PATRICIA GILAVERT FERNANDES

**Testes Adaptativos Computadorizados como um Processo de Decisão
Markoviano: equilíbrio ótimo entre eficiência e precisão**

São Paulo

2022

PATRICIA GILAVERT FERNANDES

**Testes Adaptativos Computadorizados como um Processo de Decisão
Markoviano: equilíbrio ótimo entre eficiência e precisão**

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 30 de Agosto de 2022. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Valdinei Freire da Silva

São Paulo

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada pela Biblioteca da Escola de Artes, Ciências e Humanidades,
com os dados inseridos pelo(a) autor(a)
Brenda Fontes Malheiros de Castro CRB 8-7012; Sandra Tokarevicz CRB 8-4936

Gilavert Fernandes, Patricia

Testes Adaptativos Computadorizados como um
Processo de Decisão Markoviano: equilíbrio ótimo
entre eficiência e precisão / Patricia Gilavert
Fernandes; orientador, Valdinei Freire da Silva. --
São Paulo, 2022.
66 p: il.

Dissertacao (Mestrado em Ciencias) - Programa de
Pós-Graduação em Sistemas de Informação, Escola de
Artes, Ciências e Humanidades, Universidade de São
Paulo, 2022.

Versão corrigida

1. Testes Adaptativos Computadorizados. 2.
Processos de Decisão Markovianos. 3. Critérios de
parada. I. Silva, Valdinei Freire da, orient. II.
Título.

Dissertação de autoria de Patricia Gilavert Fernandes, sob o título “**Testes Adaptativos Computadorizados como um Processo de Decisão Markoviano: equilíbrio ótimo entre eficiência e precisão**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 30 de Agosto de 2022 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Valdinei Freire da Silva
Universidade de São Paulo
Presidente

Profa. Dra. Mariana Curi
Universidade de São Paulo

Prof. Dr. Jorge Luis Bazán
Universidade de São Paulo

Dedico este trabalho à minha mãe Neide, ao meu pai Geraldo, ao meu irmão Alexandre e aos meus amigos que estiveram comigo nesse período e não soltaram a minha mão.

Agradecimentos

A Deus por ter me dado saúde, sabedoria, persistência e força para chegar até aqui diante de tantos obstáculos no meio do caminho. Obrigada por estar sempre comigo e pelo Teu grande amor.

Ao meu orientador Valdinei pela oportunidade de realizar este trabalho, pela confiança depositada em mim, pela paciência e empatia comigo em todos os momentos.

Ao meu pai Geraldo, à minha mãe Neide e ao meu irmão Alexandre por todo apoio, amor e cuidado nesse período. Amo muito vocês!

Aos meus amigos que contribuíram para que eu não desistisse desse sonho, ajudando a manter minha saúde mental diante dos desafios da vida nesse período. Em especial aos meus grandes e velhos amigos Lila, Karina, Mineiro, Nat, Denise, Gente Boa, Mito, Vaz, Nicole, Mercia e aos amigos que o Itaú me deu de presente: James, Laurinha, Luquinhas, Keylinha, Wal e Dudu, Luquitas, Pedrão, Fabão, Chun, Ale, Frans, e tantos outros. Agradeço também aos meus chefes que foram grandes incentivadores do mestrado: Tuma, Diego W e Kataguire.

À EACH USP e ao PPgSI pelo auxílio financeiro para participação e consequente publicação do artigo “Comprehensive Empirical Analysis of Stop Criteria in Computerized Adaptive Testing” no 13th International conference on computer supported education.

Resumo

GILAVERT, Patricia Fernandes. **Testes Adaptativos Computadorizados como um Processo de Decisão Markoviano: equilíbrio ótimo entre eficiência e precisão** 2022. 66 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2022.

Os Testes Adaptativos Computadorizados (Computerized Adaptive Testing - CAT) avaliam os indivíduos de forma adaptativa por meio de itens selecionados sequencialmente durante o teste, após cada resposta dada, de acordo com critérios de seleção de itens e critérios de parada do teste predefinidos. Esse tipo de avaliação se destaca em relação aos testes aplicados de forma convencional (papel e lápis) por apresentar um melhor compromisso entre precisão na estimação das habilidades e o tempo do teste. Em geral, os CATs consideram um modelo probabilístico de resposta dos indivíduos para cada item, possibilitando inferir a habilidade do indivíduo; o modelo probabilístico mais comum é o modelo de Teoria de Resposta ao Item (TRI), que permitem descrever parâmetros dos itens, como dificuldade e discriminação. Tradicionalmente, os critérios de seleção de itens utilizados nos CATs são míopes, isto é, avaliam o próximo melhor item sem considerar os itens ainda por vir em conjunto; além disso, consideram um critério de parada independente do critério de seleção de itens. Uma abordagem não míope que possibilita modelar os critérios de seleção e critérios de parada juntos são os Processos de Decisão Markovianos (Markov Decision Process - MDP). Este trabalho define um CAT utilizando MDP para obter um melhor compromisso entre a qualidade da estimativa da habilidade dos indivíduos e a quantidade de questões realizadas. Embora MDPs possibilitem solucionar o problema de forma ótima, a otimalidade de fato só é possível quando o problema é discreto e em tamanho factível. Neste trabalho analisa-se empiricamente os ganhos potenciais de modelar CATs como MDPs. Além disso, é feita uma avaliação abrangente dos critérios de parada do CAT, concluindo que o critério de Comprimento Fixo mostra uma curva de compensação de eficiência de precisão competitiva em todos os cenários. Destaca-se também que o mecanismo de estimativa e a distribuição de itens por banco influenciam o desempenho dos critérios de parada.

Palavras-chaves: Testes Adaptativos Computadorizados. Processos de Decisão Markovianos. Critérios de parada.

Abstract

GILAVERT, Patricia Fernandes. **Computerized Adaptive Testing as a Markov Decision Process: optimal balance between efficiency and precision** 2022. 66 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2022.

Computerized Adaptive Testing (CAT) adaptively assesses individuals through items selected sequentially during the test, after each given response, according to predefined item selection criteria and test stopping criteria. This type of assessment stands out in relation to tests applied in a conventional way (paper and pencil) because it presents a better compromise between precision in the estimation of skills and the test time. In general, the CATs consider a probabilistic model of the individuals' response to each item, making it possible to infer the individual's ability; the most common probabilistic model is the Item Response Theory (TRI) model, which allows describing item parameters, such as difficulty and discrimination. Traditionally, the item selection criteria used in CATs are myopic, that is, they evaluate the next best item without considering the items yet to come together; in addition, they consider a stopping criterion independent of the item selection criterion. A non-myopic approach that makes it possible to model selection criteria and stopping criteria together is the Markov Decision Process (MDP). This work defines a CAT using MDP to obtain a better compromise between the quality of the individuals' ability estimation and the quantity of questions performed. Although MDPs make it possible to solve the problem optimally, optimality is actually only possible when the problem is discrete and of a feasible size. In this work we empirically analyze the potential gains of modeling CATs as MDPs. In addition, a comprehensive assessment of the CAT stopping criteria is made, concluding that the Fixed Length criterion shows a competitive accuracy efficiency compensation curve in all scenarios. It is also noteworthy that the estimation mechanism and the distribution of items per bank influence the performance of the stopping criteria.

Keywords: Computerized Adaptive Tests. Markov Decision Processes. Stop criteria.

Lista de figuras

Figura 1 – Fluxograma de um CAT.	25
Figura 2 – Sistema modelado como MDP.	30
Figura 3 – Fluxograma da abordagem ALICAT.	35
Figura 4 – Sistema de teste de domínio sequencial com RL.	37
Figura 5 – Informações médias de bancos de 100 itens condicionadas ao traço θ e intervalo de confiança com nível de confiança $\alpha = 0,1$	42
Figura 6 – Comparação dos critérios de seleção de itens utilizando FL como critério de parada, EAP para estimativa e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.	45
Figura 7 – Comparação dos critérios de seleção de itens utilizando FL como critério de parada, ML para estimativa e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.	46
Figura 8 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens, EAP para estimativa e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.	47
Figura 9 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens, ML para estimativa e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.	48
Figura 10 – Desvio padrão de comprimento CAT para cada critério de parada usando estimativa EAP e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como o nível de deslocamento.	49
Figura 11 – Desvio padrão de comprimento CAT para cada critério de parada usando estimativa de ML e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como o nível de deslocamento.	50
Figura 12 – Diferença no comprimento do CAT em relação ao comprimento médio do CAT, dividindo a distribuição do traço em 5 faixas usando estimativa de EAP e <i>peak</i> de distribuição para parâmetro de dificuldade considerando -1 como o nível de deslocamento.	50

Figura 13 – Diferença no comprimento do CAT em relação ao comprimento médio do CAT, dividindo a distribuição do traço em 5 faixas usando estimativa de EAP e <i>peak</i> de distribuição para parâmetro de dificuldade considerando +1 como o nível de deslocamento.	51
Figura 14 – Média de todos os critérios de parada misturados usando EAP na estimativa de capacidade.	52
Figura 15 – Média de todos os critérios de parada misturados usando ML na estimativa de capacidade.	53
Figura 16 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens e ML para estimativa.	54
Figura 17 – Desvio padrão de comprimento CAT para cada critério de parada usando estimativa de ML.	54
Figura 18 – Diferença no comprimento do CAT em relação ao comprimento médio do CAT, dividindo a distribuição do traço em 5 faixas usando estimativa de ML.	55
Figura 19 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens, dividindo a distribuição do traço em 5 faixas usando estimativa de ML.	55
Figura 20 – Desvio padrão de comprimento CAT para cada critério de parada, dividindo a distribuição do traço em 5 faixas usando estimativa de ML.	56
Figura 21 – CAT enquadrado em um POMDP com estado de crença.	58
Figura 22 – Comparação de políticas baseadas em funções de pontuação míope e políticas ótimas de MDP.	61

Lista de algoritmos

Algoritmo 1 – Iteração de Valor	33
---	----

Lista de abreviaturas e siglas

ACC	Acurácia (<i>accuracy</i>)
CAT	Teste Adaptativo Computadorizado (<i>Computerized adaptive testing</i>)
CT	Mudança de Theta (<i>Change Theta</i>)
EAP	Esperança <i>a posteriori</i>
ENEM	Exame Nacional do Ensino Médio
FI	Informação de Fisher (<i>Fisher Information</i>)
FL	Comprimento Fixo (<i>Fixed Length</i>)
KL	Kullback–Leibler
KLP	Posterior Kullback–Leibler
MAP	Moda <i>a Posteriori</i>
MDP	Processo de decisão Markoviano (<i>Markov decision process</i>)
MEPV	Minimizando a Variação Posterior Esperada (<i>Minimizing the Expected Posterior Variance</i>)
MI	Máxima Informação (<i>Maximum Information</i>)
ML1	Modelo logístico com um parâmetro
ML2	Modelo logístico com dois parâmetros
ML3	Modelo logístico com três parâmetros
MMV	Método de Máxima Verossimilhança
MSE	Erro quadrático médio (<i>Mean Squared Error</i>)
MLWI	Informação Ponderadas de Máxima Verossimilhança (<i>Maximum Likelihood Weighted Information</i>)
MPWI	Informação Máxima Ponderada Posterior (<i>Maximum Posterior Weighted Information</i>)

POMDP	Processo de Decisão Markov Parcialmente Observável (<i>Partially Observable Markov Decision Process</i>)
RL	Aprendizado por reforço (<i>reinforcement learning</i>)
SE	Erro Padrão (<i>Standard Error</i>)
SMT	Teste de domínio sequencial
SSP	Caminho mais curto estocástico (<i>Stochastic Shortest Path</i>)
TRI	Teoria de Resposta ao Item
VAP	Variância <i>a Posteriori</i>
VI	Iteração de Valor (<i>Value Iteration</i>)
VVAP	Variância de Variância <i>a posteriori</i>

Lista de símbolos

θ	Escore verdadeiro. Estimativa do nível de habilidade dos indivíduos considerando a prova completa.
$\hat{\theta}$	Escore parcial. Nível de habilidade dos avaliados estimados durante a execução do CAT.
a	Parâmetro de discriminação do item
b	Parâmetro de dificuldade do item
c	Probabilidade de acerto aleatório do item
D	Constante numérica para que a função logística forneça resultados semelhantes com a da função ogiva normal
S_i	Função de pontuação
\mathcal{S}	Conjunto de estados
s	Estado do processo
\mathcal{A}	Conjunto de ações
a	Ação que pode ser executada
\mathcal{T}	Função que define a probabilidade do sistema passar de um estado para outro
\mathcal{R}	Função que dá a recompensa (ou custo)
t	Tempo
π	Política
V^π	Função valor de uma política
\mathcal{P}	Função de transição
\mathcal{C}	Função de custo
c	Custo

s_0	Estado inicial
\mathcal{G}	Conjunto de estados de meta
O	Função de observação probabilística
o	Observação
h_t	Histórico para o passo de tempo atual t
b_t	Estado de crença

Sumário

1	Introdução	17
1.1	<i>Objetivo</i>	20
1.1.1	Objetivos Específicos	20
1.2	<i>Contribuições</i>	21
1.3	<i>Organização dos capítulos</i>	21
2	Fundamentação teórica	22
2.1	<i>Modelo de resposta e estimadores de traço latente</i>	22
2.1.1	Teoria de resposta ao item - TRI	22
2.1.2	Estimadores de traço latente	23
2.2	<i>Testes adaptativos computadorizados - CAT</i>	24
2.2.1	CrITÉrios de seleção de itens	25
2.2.2	CrITÉrios de parada	27
2.2.3	Eficiência, Precisão e Restrições	28
2.3	<i>Processo de decisão Markoviano - MDP</i>	29
2.3.1	Caminho mais curto estocástico - SSP	31
2.3.2	Solução de Programação Linear	32
2.3.3	Iteração de Valor - VI	33
2.3.4	MDP Parcialmente Observável - POMDP	33
2.3.5	Estados Aumentados	34
3	Trabalhos correlatos	35
4	Avaliando critérios de parada	41
4.1	<i>Dados sintéticos</i>	41
4.1.1	Simulação CAT	42
4.1.2	Equilíbrio entre precisão e eficiência	43
4.1.3	Resultados	44
4.2	<i>Dados reais</i>	50
5	CAT como um problema de otimização	57
5.1	<i>CAT como POMDP</i>	57

5.2	<i>Aproximando um CAT por um MDP</i>	59
5.2.1	Método Bayesiano Normalizado e Discretizado	60
5.2.2	Resultados	60
6	Conclusão e considerações finais	62
	REFERÊNCIAS	64

1 Introdução

O Teste Adaptativo Computadorizado (*Computerized adaptive testing* – CAT) é uma abordagem de avaliação que adapta a administração de itens de teste ao nível de característica do examinando. Em vez de aplicar a mesma questão a todos os examinandos, como em um teste tradicional de papel e lápis, os CATs aplicam questões uma após a outra e cada seleção de questão é condicionada às questões e respostas anteriores (SEGALL, 2005). O número de questões aplicadas a cada examinando também pode variar para alcançar um melhor equilíbrio entre precisão, uma estimativa correta do traço, e eficiência, um pequeno número de questões. CATs reduzem a carga dos examinandos de duas maneiras; primeiro, os examinandos não precisam completar um longo teste; segundo, os examinandos respondem a questões adaptadas ao seu nível de característica, evitando questões muito difíceis ou muito fáceis (SPENASSATO; BORNIA; TEZZA, 2015).

Porque os examinandos não resolvem o mesmo conjunto de questões; uma estimativa apropriada do nível de traço latente do examinando deve ser considerada. No caso de questões dicotômicas, a teoria de resposta ao item (TRI) pode ser usada para encontrar a probabilidade de um examinado pontuar um item em função de seu traço θ , e, portanto, fornecer um estimador coerente. CAT em combinação com TRI torna possível calcular proficiências comparáveis entre examinandos que responderam a diferentes conjuntos de itens e em momentos diferentes (HAMBLETON; SWAMINATHAN, 2013; KREITZBERG; STOCKING; SWANSON, 1978). Essa probabilidade é influenciada por parâmetros do item, como dificuldade e discriminação.

Em cada CAT identifica-se pelo menos seis componentes (WAINER *et al.*, 2000; WANG; CHANG; HUEBNER, 2011): (i) um banco de itens, (ii) uma regra de entrada, (iii) o modelo de resposta, (iv) um mecanismo de estimativa, (v) um critério de seleção de itens e (vi) um critério de parada. O banco de itens determina as questões que estão disponíveis para a prova; geralmente, os itens são selecionados sem reposição. A regra de entrada especifica o conhecimento *a priori* do examinando; em uma estrutura Bayesiana, representa uma distribuição *a priori* sobre traços latentes e, em uma estrutura de Verossimilhança, representa uma estimativa inicial. O modelo de resposta descreve a probabilidade de pontuação de cada examinando em cada questão do banco de itens; o modelo de resposta suporta o mecanismo de estimativa para estimar o traço latente

do examinando atual. O critério de seleção de itens escolhe a questão a ser aplicada ao examinando atual, enquanto o critério de parada escolhe quando interromper o teste; geralmente, ambos os critérios podem ser suportados pela estimativa atual, pelo banco de itens e pelo modelo de resposta.

O CAT pode ser avaliado por sua precisão e eficiência e ambas as métricas dependem dos seis componentes do CAT. Muita pesquisa sobre CAT é dedicada a fornecer e avaliar diferentes critérios de seleção; enquanto os critérios de parada são menos explorados. No entanto, o critério de parada é o principal responsável por escolher o equilíbrio entre precisão e eficiência.

O Processo de Decisão de Markov (*Markov Decision Process* – MDP) modela problemas no qual um agente toma decisões sequenciais e os resultados dessas ações são probabilísticos (MAUSAM; KOLOBOV, 2012). O objetivo do agente é escolher uma sequência de ações para que o sistema atue de forma ótima de acordo com critérios previamente definidos. Essa sequência de ações segue uma política, que mapeia cada estado observado em uma ação. Enquanto os algoritmos de MDP consideram o modelo do problema de otimização conhecido, o problema de Aprendizado por reforço (*Reinforcement Learning* – RL) considera o mesmo problema de otimização quando o modelo é desconhecido e deve ser aprendido por meio de interação de tentativa e erro com o processo (SUTTON; BARTO, 1998).

Embora o CAT possa ser visto como um problema de planejamento, no qual um longo horizonte deve ser levado em consideração para atuar de forma otimizada, a maioria dos métodos CAT considera funções de pontuação míope que selecionam questões por meio de uma análise imediata. Apenas alguns trabalhos na literatura aplicaram a estrutura MDP ao CAT, principalmente enquadrando o problema de decisão sequencial como Aprendizado por Reforço. El-Alfy (2011) aplicou RL ao Teste de Domínio Sequencial, quando um CAT é projetado para classificar cada examinando como mestre ou não mestre; neste trabalho, apenas dois examinandos são considerados e as questões seguem o mesmo modelo de resposta; portanto, o estado pode ser simplesmente definido como a diferença entre a resposta errada e a certa e a única decisão é quando parar o teste. Nurakhmetov (2019) faz uso de Redes Neurais Recorrentes para modelar o espaço de estados, principalmente para dar conta de uma representação embutida de questões já aplicadas; a superexposição de itens é evitada penalizando os itens superexpostos.

Ainda que alguns trabalhos anteriores enquadrem o problema do CAT como um MDP, eles o fazem em um cenário específico. Neste trabalho mostra-se como o MDP pode modelar diferentes formalismos para CAT, em particular, os problemas de CAT sob o formalismo do Processo de Decisão Markov Parcialmente Observável (*Partially Observable Markov Decision Process* – POMDP), quando os agentes não observam o estado completamente (HOERGER; KURNIAWATI, 2020). Também é apresentado um exemplo empírico de como os MDPs podem ser usados para resolver problemas de CAT, comparando com soluções míopes tradicionais da literatura de CAT.

Além da modelagem do CAT como um MDP, foi foco também deste trabalho a avaliação abrangente dos critérios de parada. Porque o desempenho dos critérios de parada pode depender dos outros cinco componentes do CAT, uma variedade de sete critérios de parada foram testados enquanto também variou-se dois dos outros componentes: bancos de itens e mecanismo de estimativa.

Embora o modelo de resposta e a regra de entrada possam influenciar o desempenho dos critérios de parada, ambos apresentam escolhas naturais na prática. Para o modelo de resposta, por conta dos itens dicotômicos, optou-se pelo modelo logístico de três parâmetros. Para a regra de entrada, considera-se a distribuição normal padrão, que é comumente utilizada para calibrar os bancos de itens. Embora os critérios de seleção apresentem muitas opções, todos apresentam o mesmo comportamento: quanto maior o número de questões, melhor a precisão. Os critérios de parada visam equilibrar o nível de precisão e a possibilidade de melhoria para a população de examinandos. Avaliou-se os critérios de parada fixando o critério de seleção com o critério Informação de Fisher (*Fisher Information* – FI) (LORD, 1980). O critério FI é amplamente utilizado por ser o mais barato computacionalmente.

Para avaliar os critérios de parada, neste trabalho é proposto um método de compensação de precisão e eficiência. A maioria dos critérios de parada considera uma métrica e um limiar; se a métrica estiver abaixo do limite, o teste será encerrado. Normalmente, os trabalhos que avaliam os critérios de parada consideram um pequeno conjunto de configurações para cada critério de parada e medem a eficiência e a precisão de cada configuração; porque nem eficiência nem precisão são fixas, verifica-se que tais configurações são incomparáveis. O método proposto considera a configuração para obter níveis de eficiência ao longo de todo o espectro do número de questões (eficiência) e precisão, resultando em uma curva de compensação entre precisão e eficiência para cada critério de parada.

Essa curva de compensação permite comparar os critérios de parada ao longo do espectro relevante de limites.

Outra proposta é um experimento considerando simulação sobre uma grande quantidade de bancos de itens sintéticos. Primeiro, considera-se diferentes configurações para gerar bancos de itens; varia-se em três classes de distribuição de probabilidade para dificuldade de parâmetro dos itens e em três níveis de centralidade para cada classe. Segundo, para cada configuração, 500 bancos de itens são simulados. Trabalhos na literatura consideram apenas um banco de itens para cada configuração o que pode aumentar o viés. Apesar de não ser o foco deste trabalho, aproveita-se essa miríade de bancos de itens para também avaliar diferentes critérios de seleção sob o critério de parada de Comprimento Fixo. O experimento foi replicado para dados reais, usando a prova do ENEM de 2012, para comparar os resultados.

O pequeno número de trabalho na literatura avaliando critérios de parada justifica as propostas apresentadas. O trabalho de [Babcock e Weiss \(2009\)](#) é uma inspiração para selecionar diferentes classes de distribuição de probabilidade para dificuldade de item. Eles fazem uso de duas classes (uniforme e pico) e dois comprimentos de bancos de itens (100 e 500). No entanto, eles simulam apenas uma instância de cada configuração do banco de itens. O trabalho de [Morris *et al.* \(2020\)](#) é uma inspiração para selecionar diferentes centralidades para dificuldade dos itens. Eles experimentam um banco de itens real para avaliar os resultados relatados pelo paciente; tal banco de itens tem uma centralidade positiva sobre a dificuldade do item.

1.1 *Objetivo*

O objetivo deste trabalho é demonstrar como MDPs podem ser utilizados para modelar CATs em geral e avaliar qual o ganho que tal modelagem pode agregar em CATs tradicionais (questões monotônicas, uma questão por vez sem reposição, e avaliação de traço latente).

1.1.1 *Objetivos Específicos*

Este trabalho apresenta os seguintes objetivos específicos:

- Propor um método de avaliação para diferentes critério de parada;
- Fazer uma avaliação abrangente dos critérios de parada da literatura;
- Propor um CAT irreal, mas que possa ser resolvido de forma ótima por algoritmos de MDPs; e
- Fazer comparação dos resultados com dados simulados e reais.

1.2 Contribuições

Em resumo, as contribuições deste trabalho são:

- Modelagem de problemas de CAT com MDPs;
- Método de avaliação que permite comparar equilíbrios entre precisão e eficiência para critérios de parada;
- Avaliação abrangente de critérios de parada da literatura em dados sintéticos e reais;
- Especificação de um CAT irreal cuja solução ótima exista; e
- Avaliação de abordagem com MDP comparada a métodos tradicionais da literatura.

1.3 Organização dos capítulos

O capítulo 2 introduz os conceitos teóricos principais de CAT e MDP. O capítulo 3 mostra os trabalhos correlatos. A avaliação dos critérios de parada é apresentada no capítulo 4. A construção do CAT como MDP é descrita no capítulo 5 e finalmente, no capítulo 6 são apresentadas as conclusões e considerações finais.

2 Fundamentação teórica

Neste capítulo serão detalhados os fundamentos teóricos discutidos no presente trabalho. Na seção 2.1 é apresentado o modelo de resposta e os estimadores de traço latente. Na seção 2.2 e 2.3 são detalhados os testes adaptativos computadorizados e o processo de decisão Markoviano, respectivamente.

2.1 Modelo de resposta e estimadores de traço latente

Nesta seção é apresentado o modelo de resposta e os estimadores de traço latente utilizados no trabalho.

2.1.1 Teoria de resposta ao item - TRI

A Teoria da Resposta ao Item (TRI) trata de modelos matemáticos que procuram descrever a probabilidade de um indivíduo obter um certo escore em um item em função do nível do traço latente (ou dos traços latentes). Essa probabilidade também é influenciada por parâmetros do item, como dificuldade e capacidade de discriminação.

Os modelos ML1 (modelo logístico de um parâmetro), ML2 (modelo logístico de dois parâmetros) e ML3 (modelo logístico de três parâmetros) podem ser utilizados para o cálculo da probabilidade de acerto para dados com respostas dicotômicas (ou seja, certo ou errado). O ML1, também chamado de modelo Rasch, possui uma dimensão que considera apenas o parâmetro de dificuldade do item (RASCH, 1960). O ML2 é o modelo logístico unidimensional com os parâmetros de dificuldade e discriminação (BAKER; KIM, 2004). O ML3 é o modelo com três parâmetros, dificuldade, discriminação e probabilidade de acerto ao acaso (BIRNBAUM, 1968), dado por:

$$\Pr(X_i = 1 | \theta) = c_i + \frac{(1 - c_i)}{1 + \exp[-Da_i(\theta - b_j)]}, \quad (1)$$

onde,

- θ é o traço latente do examinando, no caso de um teste, θ é a habilidade do examinando;
- X_i é uma variável binária, 1 indica que o examinando responde corretamente ao item i e 0 caso contrário;

- $\Pr(X_i = 1 | \theta)$ é a probabilidade de um indivíduo com traço latente θ responder corretamente ao item i ;
- a_i é o parâmetro de discriminação do item i ;
- b_i é o parâmetro de dificuldade do item i ;
- c_i é a probabilidade de uma resposta correta aleatória do item i ; e
- D é um fator de escala que equivale a 1 para o modelo logístico e 1,702 para que a função logística se aproxime da função ogiva normal.

Os modelos ML1 e ML2 são casos particulares do modelo ML3. Assim, quando $c_i = 0$ chega-se no modelo ML2 e quando, além de $c_i = 0$, o parâmetro de discriminação a_i for igual para todos os itens chega-se no modelo ML1.

2.1.2 Estimadores de traço latente

Genericamente, pode-se considerar um traço latente $\theta_j \in \Theta$ para um examinando j , onde Θ é o conjunto de suporte para o traço latente. Essa característica pode ser unidimensional ou multidimensional. Quando uma questão i é submetida ao examinando θ_j , uma resposta é dada e um resultado x_{ij} é observado.

Os resultados podem ser dicotômicos, politômicos ou contínuos. Um modelo de resposta considera questões parametrizadas, ou seja, uma questão i é representada por um vetor de parâmetros γ_i e um conjunto de suporte \mathcal{X}_i para possíveis resultados. Após uma questão i ser submetida a um examinando j , uma variável aleatória $X_{ij} \in \mathcal{X}_i$ representando o resultado observado é gerada e o modelo de resposta define uma distribuição de probabilidade para os resultados possíveis¹, ou seja, para todo $x \in \mathcal{X}_i$:

$$\Pr(X_{ij} = x | \gamma_i, \theta_j).$$

Dado um examinado e uma sequência de n respostas $\mathbf{x}_n = x_{i_1}, x_{i_2}, \dots, x_{i_n}$, o traço latente θ pode ser estimado por procedimento Bayesiano ou Máxima Verossimilhança (ML) (ANDRADE; TAVARES; VALLE, 2000).

Considerou-se aqui um estimador Bayesiano baseado na esperança *a posteriori* (EAP), ou seja,

$$\hat{\theta} = E[\theta | \mathbf{x}_n] = \int \theta \frac{f(\theta) \prod_{k=1}^n \Pr(X_{i_k} = x_{i_k} | \theta)}{\Pr(\mathbf{X}_n = \mathbf{x}_n)} d(\theta), \quad (2)$$

¹ Considerou-se o caso em que \mathcal{X}_i é enumerado para uma exposição mais breve.

onde $f(\theta)$ é a distribuição a priori do traço latente θ , normalmente considerada a distribuição normal padrão.

O estimador ML estima o traço latente por $\hat{\theta} = \max_{\theta} L(\theta|\mathbf{x}_n)$ onde a verossimilhança é dada por:

$$L(\theta|\mathbf{x}_n) = \prod_{k=1}^n \Pr(X_{i_k} = x_{i_k}|\theta). \quad (3)$$

Se o objetivo do CAT é classificar um examinando entre a categoria de opções \mathcal{C} , uma função de classificação $C : \Theta \rightarrow \mathcal{C}$ pode ser definida após a estimativa do traço latente $\hat{\theta}$ ou um estimador de categoria \hat{C} pode ser definido diretamente.

2.2 Testes adaptativos computadorizados - CAT

Normalmente, um CAT é usado para estimar algum traço latente de um examinando. Uma universidade pode usar o CAT para classificar o aluno estimando uma nota para cada candidato. Um médico pode diagnosticar alguma condição do paciente sob um espectro multidimensional. O governo pode dar uma carteira de motorista para um adolescente.

Dada a existência de uma metodologia consolidada e efetiva como a TRI para a elaboração de provas mais eficazes e justas com o desempenho do estudante, é possível construir um CAT baseado na TRI.

Os CATs são aplicados de forma adaptativa a cada examinando por computador. Com base nas regras pré-definidas do algoritmo, os itens são selecionados sequencialmente durante o teste após cada resposta a um item (SPENASSATO; BORNIA; TEZZA, 2015). Um CAT clássico pode ser descrito pelas seguintes etapas (LINDEN; GLAS, 2000):

1. O primeiro item é selecionado;
2. O traço latente é estimado com base na resposta do primeiro item;
3. O próximo item a ser respondido é selecionado;
4. O traço latente é recalculado com base nas respostas anteriores; e
5. Repita os passos 3 e 4 até que uma resposta não seja mais necessária de acordo com um critério pré-estabelecido, chamado critério de parada.

Em geral, o fluxograma do CAT é como na Figura 1.

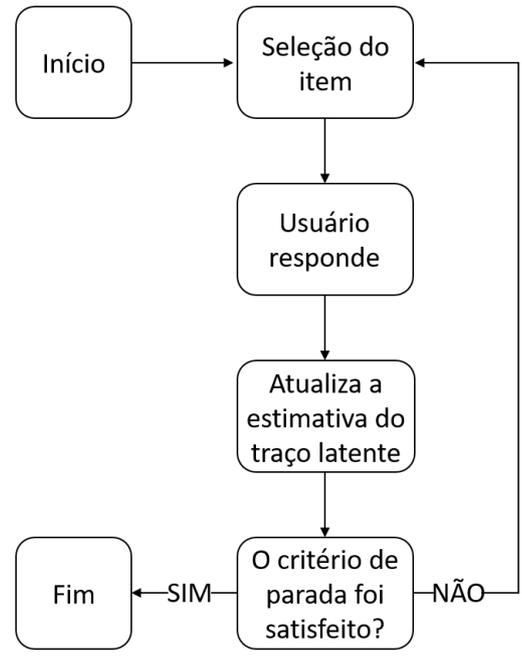


Figura 1 – Fluxograma de um CAT.

Fonte – Patricia Gilavert, 2020

2.2.1 Critérios de seleção de itens

A escolha do método de seleção de itens pode ter efeito na eficiência e precisão da estimativa da habilidade do examinando. Neste trabalho, foram considerados cinco critérios de seleção de itens diferentes. Três deles são baseados em Informação de Fisher, enquanto dois deles são baseados na divergência de Kullback-Leibler.

Cada critério define uma função de pontuação $S_i(\mathbf{x}_n)$ para cada item i dadas n respostas anteriores do examinando, então, entre os itens que ainda não foram aplicados ao examinando, aquele com a maior pontuação é escolhido.

Em cada etapa $n + 1$, ao selecionar um item, os critérios de seleção de itens podem fazer uso de: $\hat{\theta}_n$, a estimativa do traço latente após n respostas; $f(\theta|\mathbf{x}_n)$, a distribuição *a posteriori* após n respostas; e $L(\theta|\mathbf{x}_n)$, a verossimilhança após n respostas. Para simplificar a notação, descreve-se brevemente $P_i(\theta) = \Pr(X_i = 1 | \theta)$.

Também definiu-se a divergência de Kullback-Leibler entre a distribuição de pontuação do item i para dois examinandos com traço latente θ e $\hat{\theta}$ por

$$KL_i(\theta||\hat{\theta}) = P_i(\hat{\theta}) \ln \left[\frac{P_i(\hat{\theta})}{P_i(\theta)} \right] + Q_i(\hat{\theta}) \ln \left[\frac{Q_i(\hat{\theta})}{Q_i(\theta)} \right] \quad (4)$$

onde $Q_i(\theta) = 1 - P_i(\theta)$.

Informação de Fisher (*Fisher Information* – FI) (LORD, 1980): este método seleciona o próximo item que maximiza a informação de Fisher dada a estimativa do traço latente (SARI; RABORN, 2018), ou seja,

$$S_i(\mathbf{x}_n) = I_i(\hat{\theta}_n) = \frac{\left[\frac{d}{d\hat{\theta}_n} P_i(\hat{\theta}_n) \right]^2}{P_i(\hat{\theta}_n)(1 - P_i(\hat{\theta}_n))} \quad (5)$$

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

Kullback–Leibler (KL) (CHANG; YING, 1996): é baseado em um teste de razão de verossimilhança. Na estrutura CAT, esse método calcula a divergência não simétrica entre duas probabilidades em dois níveis estimados de característica, chamado ganho de informação KL. KL é a razão de duas funções de verossimilhança em vez de um valor fixo como no FI (SARI; RABORN, 2018). O critério KL define a seguinte função de pontuação:

$$S_i(\mathbf{x}_n) = \int_{-\infty}^{\infty} KL_i(\theta|\hat{\theta}_n)L(\theta|\mathbf{x}_n)d(\theta). \quad (6)$$

Posterior Kullback–Leibler (KLP) (CHANG; YING, 1996): o método KLP pondera as informações atuais de KL pela distribuição *a posteriori* de θ (SARI; RABORN, 2018). O critério KLP define a seguinte função de pontuação:

$$S_i(\mathbf{x}_n) = \int_{-\infty}^{\infty} KL_i(\theta|\hat{\theta}_n)f(\theta|\mathbf{x}_n)d(\theta). \quad (7)$$

Informação Ponderada de Máxima Verossimilhança (*Maximum Likelihood Weighted Information* – MLWI) (VEERKAMP; BERGER, 1997): enquanto FI considera as informações de Fisher na estimativa atual $\hat{\theta}_n$, MLWI pondera informações de Fisher em diferentes níveis pela função de verossimilhança (SARI; RABORN, 2018), ou seja,

$$S_i(\mathbf{x}_n) = \int_{-\infty}^{\infty} I_i(\theta)L(\theta|\mathbf{x}_n)d(\theta). \quad (8)$$

Informação Máxima Ponderada Posterior (*Maximum Posterior Weighted Information* – MPWI) (LINDEN, 1998): assim como MLWI, MPWI considera uma informação de Fisher de peso, mas neste caso considerando a distribuição *a posteriori* (SARI; RABORN, 2018), ou seja,

$$S_i(\mathbf{x}_n) = \int_{-\infty}^{\infty} I_i(\theta)f(\theta|\mathbf{x}_n)d(\theta). \quad (9)$$

Minimizando a Variação Posterior Esperada (*Minimizing the Expected Posterior Variance* – MEPV) (MORRIS *et al.*, 2020): A abordagem Bayesiana é

considerada. O CAT apresenta o item i para o qual o valor esperado da variância posterior, dado que administrou-se o item i , é menor, ou seja,

$$S_i(\mathbf{x}_n) = -E_i(\text{Var}(\theta|\mathbf{x}_n)). \quad (10)$$

2.2.2 Critérios de parada

Enquanto os Critérios de Seleção de Itens têm o objetivo de determinar um bom compromisso entre a duração do teste e a estimativa de característica latente, os Critérios de Parada elegem o melhor compromisso para um determinado teste. Nesse caso, um agente de teste deve determinar a importância da duração do teste, eficiência, e da qualidade da estimativa, precisão.

Comprimento Fixo (*Fixed Length* – FL): é o critério de parada mais comum. Nesse caso, cada examinado responde a um subconjunto das questões N , potencialmente para um subconjunto diferente de itens. Enquanto FL garante que todos os examinados respondam ao mesmo número de questões, proporcionando alguma sensação de justiça, os examinados podem ser avaliados com precisão diferente, a menos que o número de questões respondidas seja grande o suficiente.

Os critérios de parada de comprimento variável podem ser agrupados em dois grupos: **Precisão Mínima** e **Informações Mínimas**. A primeira interrompe um teste somente quando se obtém uma precisão mínima na estimativa do traço latente. A segunda interrompe um teste se não houver mais informações nos bancos de itens. Os critérios de parada diferenciam-se em como a precisão e a informação são medidas.

Erro Padrão (*Squared Error* – SE) (BABCOCK; WEISS, 2009): considera a precisão dada pelo desvio padrão do estimador de traço latente $\hat{\theta}_n$. Se a característica latente real θ_0 for conhecida, o desvio padrão pode ser calculado pela informação de Fisher; ou seja,

$$\sqrt{\text{Var}(\hat{\theta}_n)} = \frac{1}{\sqrt{\frac{\partial^2 \log L(\theta_0|\mathbf{x}_n)}{\partial \theta^2}}}.$$

Como θ_0 é desconhecido, é aproximado por $\hat{\theta}_n$ e o erro padrão é definido por:

$$SE(\hat{\theta}_n) = \frac{1}{\sqrt{\frac{\partial^2 \log L(\hat{\theta}_n|\mathbf{x}_n)}{\partial \theta^2}}}. \quad (11)$$

Variância *a Posteriori* (VAP): semelhante ao SE, quando uma abordagem Bayesiana é considerada, uma precisão sobre estimativa pode ser obtida pela variância da distribuição *a posteriori* $f(\theta|\mathbf{x}_n)$. Portanto, define-se:

$$VAP(\hat{\theta}_n) = \text{Var}(\theta) = \int (\theta - \hat{\theta}_n)^2 f(\theta|\mathbf{x}_n) d(\theta). \quad (12)$$

Máxima Informação (*Maximum Information* – MI) (BABCOCK; WEISS, 2009): considera as informações para cada questão ainda não submetida ao examinando. A intuição é que se nenhuma questão tiver informação, então, o teste pode parar. Portanto, define-se:

$$MI(\hat{\theta}_n) = \max_{i \in \mathcal{Q}_n} I_i(\hat{\theta}_n), \quad (13)$$

onde \mathcal{Q}_n é o conjunto de itens não submetidos ao examinando no estágio n .

Mudança de Theta (*Change Theta* – CT) (STAFFORD *et al.*, 2019): enquanto o MI avalia as questões antes de enviá-las a um examinando, o critério de parada CT avalia as informações da última questão pela quantidade de mudança no estimador $\hat{\theta}_n$, ou seja,

$$CT(\hat{\theta}_n) = |\hat{\theta}_n - \hat{\theta}_{n-1}|. \quad (14)$$

Variância de Variância *a Posteriori* (VVAP): um novo critério de parada baseado na variância *a posteriori* é proposto. O objetivo é comparar as variâncias *a posteriori* dos θ calculados nos últimos itens administrados. Caso não esteja mudando, o teste pode ser interrompido porque não há mais informações. Considerou-se a variância da variância *a posteriori* das 5 últimas estimativas e definiu-se:

$$VVAP(\hat{\theta}_n) = \sum_{j=0}^4 \frac{(\text{Var}(\hat{\theta}_{n-j}) - \mu_{Var,n})^2}{5}, \quad (15)$$

onde

$$\mu_{Var,n} = \sum_{j=0}^4 \frac{\text{Var}(\hat{\theta}_{n-j})}{5}.$$

2.2.3 Eficiência, Precisão e Restrições

Um CAT pode ser avaliado sob dois critérios principais: precisão e eficiência. A precisão está relacionada ao quão boa é a estimativa $\hat{\theta}$, enquanto a eficiência está relacionada ao esforço feito pelo examinando ou examinador.

Se o objetivo do CAT é classificar os examinandos, uma avaliação comum para precisão é o erro quadrático médio (*Mean Squared Error* – MSE):

$$MSE = E[(\theta - \hat{\theta})^2] = \int (\theta - \hat{\theta})^2 f(\theta) d(\theta).$$

Se o objetivo do CAT é classificar um examinando, uma avaliação comum para precisão pode ser acurácia (*Accuracy* – ACC):

$$ACC = E[1_{C_\theta = \hat{c}}],$$

onde 1_A é a função indicadora para a condição A e C_θ é a categoria correta.

A eficiência geralmente é avaliada pela duração do CAT, ou seja, quantas questões o examinando respondeu. A eficiência também pode ser avaliada pelo tempo gasto pelo examinando, pois algumas questões podem exigir mais tempo do que outras, ou alguma outra medida de esforço.

Finalmente, o CAT pode buscar precisão e eficiência sob alguma restrição. A restrição mais comum é quanto à repetição de questões; em um teste uma questão só pode ser aplicada uma vez. Um examinador também pode se preocupar com o balanceamento do conteúdo do teste e o controle da exposição do item. O balanceamento de conteúdo considera que o banco de itens está agrupado em grupos de questões e o teste deve escolher questões de todos os grupos com uma taxa mínima. O controle de exposição do item considera que uma questão não deve ser submetida a muitos examinandos; se uma questão for superexposta, os examinandos saberão de antemão e o CAT perderá a precisão.

Uma restrição específica é considerada no CAT multiestágio (MAGIS; YAN; DAVIER, 2017). No CAT multiestágio, o teste é subdividido em etapas e, em qualquer etapa, um conjunto de questões deve ser revelado de uma só vez ao examinando. Depois que o examinando responde a todas as questões em uma etapa, ele é encaminhado individualmente para uma nova etapa.

2.3 Processo de decisão Markoviado - MDP

Como mencionado, os critérios de seleção de itens utilizados nos CATs clássicos são míopes, ou seja, avaliam o próximo melhor item sem considerar os próximos itens, e o critério de parada é fixo. Uma forma de não ser míope e modelar os critérios de seleção e critérios de parada juntos é tratando o problema como um MDP.

Um MDP é um modelo de um agente que interage em sincronia com um ambiente. O agente recebe como entrada o estado do ambiente e gera como saída ações, que por sua vez afetam o estado do ambiente. Na estrutura do MDP, presume-se que, embora possa haver uma grande incerteza sobre os efeitos das ações de um agente, nunca há qualquer incerteza sobre o estado atual do agente, por ter habilidades perceptivas completas e perfeitas (KAELBLING; LITTMAN; CASSANDRA, 1998).

O MDP pode ser descrito por uma tupla $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$, onde:

- \mathcal{S} é um conjunto de estados em que o processo pode estar;
- \mathcal{A} é um conjunto de ações que podem ser executadas em diferentes épocas de decisão;
- $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \mapsto [0, 1]$ é uma função que define a probabilidade do sistema passar para um estado $s' \in \mathcal{S}$, dado que o processo estava em um estado $s \in \mathcal{S}$ e o agente decidiu executar uma ação $a \in \mathcal{A}$ ($\mathcal{T}(s'|s, a)$);
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{R}$ é uma função que dá a recompensa (ou custo) por tomar uma decisão $a \in \mathcal{A}$ quando o processo está em um estado $s \in \mathcal{S}$.

Na propriedade de Markov, o próximo estado e a recompensa esperada dependem apenas do estado anterior e da ação realizada. O estado e a recompensa no tempo $t + 1$ dependem apenas do estado no tempo t e da ação no tempo t .

A Figura 2 mostra um sistema modelado como um MDP. O tomador de decisões verifica o estado atual do sistema (s), consulta uma política (π) e executa uma ação (a). Esta ação pode ter um efeito sobre o ambiente e modificar o estado atual. O tomador de decisões, então, verifica o novo estado para que possa tomar a próxima decisão.

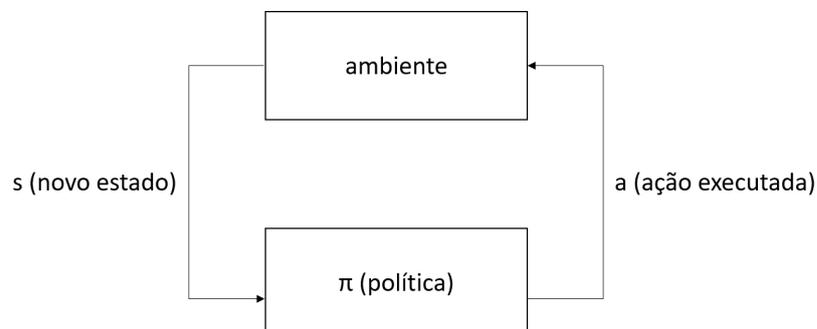


Figura 2 – Sistema modelado como MDP.

Fonte – Adaptado de Pellegrini e Wainer (2007).

A cada época de decisão, o tomador de decisões usa uma regra de decisão para escolher a próxima ação. O conjunto de todas as regras de decisão (uma para cada época

de decisão) é chamado de política π . O MDP tem como objetivo encontrar a política ideal que otimiza uma determinada métrica de desempenho, resultando em maiores recompensas dado cada estado.

Uma recompensa imediata, que pode ser positiva, negativa ou zero, é obtida na transição de um estado para outro sob a influência de uma ação. A métrica de desempenho de uma política é geralmente uma função (função objetivo) das recompensas imediatas obtidas quando a política associada é seguida ao longo de um horizonte de tempo predeterminado. O horizonte de tempo pode ser finito ou infinito (GOSAVI, 2009).

A função valor de uma política π para um MDP é uma função $V^\pi : S \mapsto \mathbb{R}$, tal que $V^\pi(S)$ fornece o valor esperado da recompensa para a política, de acordo com um critério de desempenho (ou de “otimalidade”).

2.3.1 Caminho mais curto estocástico - SSP

Um caso especial do MDP é o problema do Caminho mais curto estocástico (*Stochastic Shortest Path* – SSP) (BERTSEKAS; TSITSIKLIS, 1991). Um SSP é definido por uma tupla $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C}, s_0, \mathcal{G} \rangle$ onde: $s \in \mathcal{S}$ são os estados possíveis; $a \in \mathcal{A}$ são ações possíveis; $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ é a função de transição; $\mathcal{C} : \mathcal{S} \times \mathcal{A} \rightarrow R^+$ é a função de custo; $s_0 \in \mathcal{S}$ é o estado inicial; e \mathcal{G} é o conjunto de estados de meta.

Um SSP define um processo de um agente interagindo com um ambiente e em todos os passos de tempo t : (i) o agente observa um estado s_t , (ii) o agente escolhe uma ação a_t , (iii) o agente paga um custo c_t ; e (iv) o processo se move para um novo estado s_{t+1} . O processo termina quando um estado objetivo $s \in \mathcal{G}$ é alcançado. Transições e custos apresentam a propriedade de Markov, ou seja, ambos dependem apenas do estado atual s_t e da ação escolhida a_t .

A solução para um SSP consiste em políticas que descrevem quais ações devem ser tomadas em cada situação. Aqui considera-se políticas estacionárias probabilísticas $\pi : \mathcal{S} \rightarrow (\mathcal{A} \rightarrow [0, 1])$ que mapeia cada estado para uma distribuição de probabilidade sobre ações.

O objetivo de um SSP é encontrar uma política ótima π^* que minimize o custo acumulado esperado, ou seja, define-se uma função valor de uma política π por $V^\pi = E[\sum_{t=0}^{\infty} c_t | \pi]$ e define-se a política ideal como $\pi^* = \arg \max_{\pi} V^\pi$. Existem muitos algoritmos

na literatura para encontrar políticas ótimas para MDPs (MAUSAM; KOLOBOV, 2012). Neste trabalho, o foco foi em algoritmos de Iteração de Valor, mas a formulação baseado em algoritmos de Programação Linear também é importante quando se deseja impor restrições (TREVIZAN; TEICHTIL-KÖNIGSBUCH; THIÉBAUX, 2017).

2.3.2 Solução de Programação Linear

A Formulação Dual de Programação Linear considera as variáveis $x_{s,a}$ para todos os $s \in \mathcal{S}, a \in \mathcal{A}$ que indicam uma frequência de ocorrência acumulada esperada para cada par estado-ação. A dinâmica SSP restringe as soluções especificando um modelo de fluxo $in(s)$ e $out(s)$ para cada estado s . Todos os estados, exceto o estado inicial e o estado da meta, devem igualar $in(s)$ e $out(s)$. O estado inicial s_0 apresenta uma diferença unitária de entrada e saída, enquanto os estados de meta em \mathcal{G} não têm saída. Para cada estado $s \in \mathcal{S}$, tem-se:

$$in(s) = \sum_{s' \in \mathcal{S}, a \in \mathcal{A}} x_{s',a} \mathcal{P}(s|s',a) \quad \text{and} \quad out(s) = \sum_{a \in \mathcal{A}(s)} x_{s,a}. \quad (16)$$

Define-se a programação linear LP1 da seguinte forma:

$$\begin{aligned} \text{LP1} \quad & \min_{x_{s,a}} \sum_{s \in \mathcal{S}, a \in \mathcal{A}} x_{s,a} \mathcal{C}(s,a) \\ & \text{s.t.} \quad x_{s,a} \geq 0 \quad \forall s \in \mathcal{S}, a \in \mathcal{A}(s) \\ & \quad \quad out(s) - in(s) \leq 0 \quad \forall s \in \mathcal{S} \setminus (\mathcal{G} \cup s_0) . \\ & \quad \quad out(s_0) - in(s_0) \leq 1 \\ & \quad \quad \sum_{s_g \in \mathcal{G}} in(s_g) = 1 \end{aligned} \quad (17)$$

A política ótima π^* para o SSP pode ser obtida da frequência de ocorrência acumulada esperada $x_{s,a}$ no LP1 por:

$$\pi^*(a|s) = \frac{x_{s,a}}{\sum_{a' \in \mathcal{A}} x_{s,a'}} \quad \forall s \in \mathcal{S},$$

e π^* incorre no custo esperado $c_{min} = \sum_{s \in \mathcal{S}, a \in \mathcal{A}} x_{s,a} \mathcal{C}(s,a)$.

2.3.3 Iteração de Valor - VI

O algoritmo de Iteração de Valor (*Value Iteration* – VI) é um algoritmo de programação dinâmica. Em cada iteração i é calculado o valor $V^i(S)$ baseado no valor $V^{i-1}(S)$ para cada estado $s \in \mathcal{S}$ do MDP, isto é:

$$V^i(s) \leftarrow \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} \mathcal{T}(s'|s, a) [\mathcal{R}(s, a) + V^{i-1}(s')], \forall s \in \mathcal{S}. \quad (18)$$

Ao longo das iterações o valor de V^i converge para a função valor ótima V^* , mesmo partindo de qualquer estimativa inicial V^0 .

Se um horizonte finito H é considerado, a função valor deve considerar o tempo como parâmetro e a função valor pode ser obtida em H iterações com:

$$V(s, t) \leftarrow \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} \mathcal{T}(s'|s, a) [\mathcal{R}(s, a) + V(s', t + 1)], \forall s \in \mathcal{S}, \quad (19)$$

e $V(s, H) = 0$ para todo $s \in \mathcal{S}$.

O Algoritmo 1 recebe como entrada um MDP e um horizonte H que é usado no critério de parada; e devolve como saída o valor ótimo para todo estado e tempo do problema (PUTERMAN, 2014).

Algoritmo 1 Iteração de Valor

Entrada: MDP $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$, H

Saída: Função valor V para o MDP de entrada

```

1 para cada  $s \in \mathcal{S}$  faça
2   |  $V(s, H) \leftarrow 0$ 
3 fim
4  $i \leftarrow 1$ 
5 para cada  $t \in \{H - 1, H - 2, \dots, 1, 0\}$  faça
6   | para cada  $s \in \mathcal{S}$  faça
7     |  $V(s, t) \leftarrow \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} \mathcal{T}(s'|s, a) [\mathcal{R}(s, a) + V(s', t + 1)]$ 
8     fim
9   fim
10 retorna  $V$ 

```

Fonte – Adaptado de Puterman (2014)

2.3.4 MDP Parcialmente Observável - POMDP

Uma formulação para o problema de decisão sequencial mais genérica que os MDPs ou SSPs é o MDP Parcialmente Observável (Markov Decision Process Parcial-Observável

POMDP) (HOERGER; KURNIAWATI, 2020). Em um MDP, considera-se que o agente observa imediatamente o estado do processo, o que garante que uma política ótima possa ser encontrada no espaço de estados, ou seja, a decisão ótima é baseada apenas na observação do estado atual.

Em um POMDP, o agente observa o estado do processo mediado por uma função de observação probabilística $O : \mathcal{S} \rightarrow (\mathcal{O} \rightarrow [0, 1])$. Um POMDP define um processo de um agente interagindo com um ambiente e em todo tempo t : (i) o agente faz uma observação $o_t \sim O(s_t)$, (ii) o agente escolhe uma ação a_t , (iii) o agente paga um custo c_t ; e (iv) o processo se move para um novo estado s_{t+1} .

Em um POMDP, a política ótima deve considerar o histórico para o passo de tempo atual t , ou seja, $h_t = (o_0, a_0, c_0, \dots, o_t)$ ou um estado de crença b_t que é uma distribuição de probabilidade sobre o espaço de estados dado o histórico de observações. Observe que o histórico cresce exponencialmente com o passo de tempo t e encontrar uma política ótima para um POMDP é apenas computacionalmente prático quando o espaço de estados é finito e pequeno. Normalmente, os algoritmos tentam encontrar uma política quase ótima (HOERGER; KURNIAWATI, 2020).

2.3.5 Estados Aumentados

MDPs e SSPs são muito estritos no que diz respeito à propriedade de Markov. Primeiro, o agente deve observar o estado completo do processo. Em segundo lugar, a função de custo deve depender apenas do estado atual. Terceiro, a função objetivo considera uma variável aleatória que é simplesmente a soma dos custos imediatos. Se alguma dessas propriedades não for apresentada, na maioria das vezes, a política ótima não se restringe ao espaço de estados do MDP.

A solução POMDP é um exemplo. Nesse caso, a política deve ser definida em um espaço de estado aumentado, o espaço de histórico. Em geral, a propriedade de Markov pode ser recuperada sob um espaço de estados aumentado apropriado. Por exemplo, se o processo terminar após algum passo de tempo fixo T , o espaço de estados deve ser aumentado com o tempo atual t , ou seja, (s, t) . Embora o espaço de estados aumentado possa ser uma técnica geral para descobrir políticas ótimas, com o POMDP o espaço de estados aumentado pode crescer exponencialmente.

3 Trabalhos correlatos

Este capítulo trata de estudos encontrados na literatura com abordagens semelhantes as propostas aqui e com resultados promissores. Esses trabalhos motivam o desenvolvimento da pesquisa envolvendo CAT e MDP.

O trabalho em [Jatobá et al. \(2020\)](#) cria uma abordagem que personaliza o processo de seleção de itens no CAT, considerando o uso de mais de um critério para escolha do próximo item. Após a análise do desempenho de diferentes critérios por meio de um estudo de caso, chega-se a conclusão de que a regra de seleção de Kullback-Leibler com distribuição *a posteriori* (KLP) possui o melhor desempenho na estimativa das habilidades em relação a outras regras testadas. Observou-se também que as regras de Informação de Fisher (F) e Informação Ponderada pela Máxima Verossimilhança (MLWI) tiveram melhor desempenho nos estágios iniciais do CAT, para estimar respondentes com níveis de habilidades extremos negativos e positivos, respectivamente. O critério de parada foi definido com base no ponto de estabilidade, ou seja, o momento do teste no qual a diferença do erro padrão entre o item atual aplicado e o item anterior é inferior a 1% do erro padrão do item anterior. A Figura 3 mostra o fluxograma da abordagem construída.

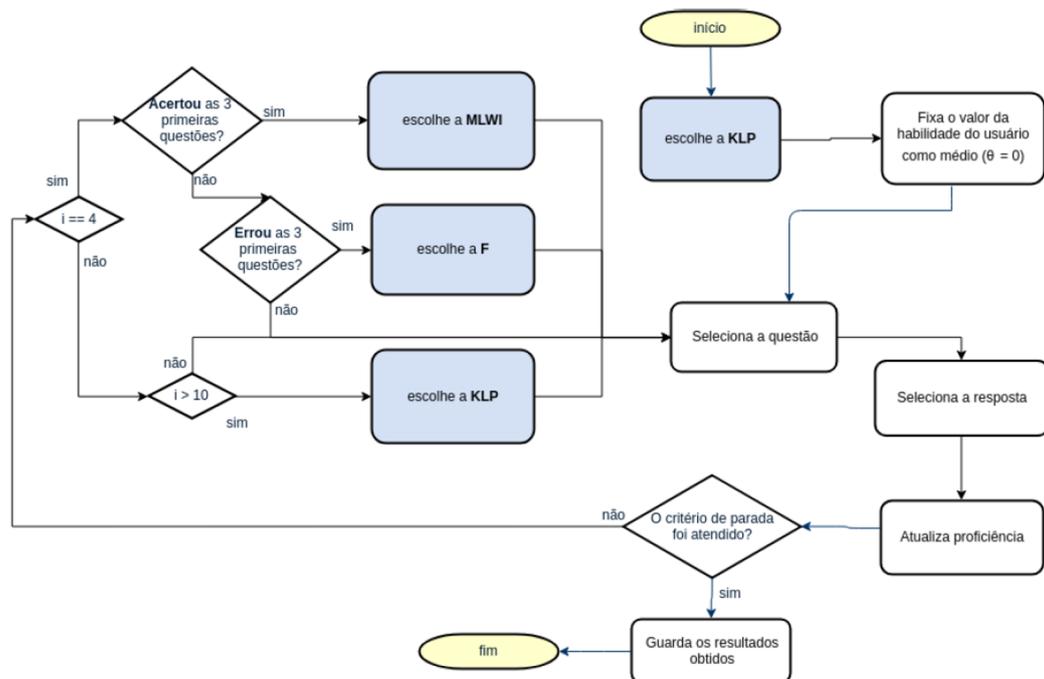


Figura 3 – Fluxograma da abordagem ALICAT.

Fonte – Adaptado de [Jatobá et al. \(2020\)](#)

No início do teste utiliza-se o método KLP, que mostrou ter o melhor desempenho na estimativa das habilidades. Após as 3 primeiras questões, utiliza-se o critério MLWI se o indivíduo acerta todas, já que o critério apresentou melhor desempenho para os extremos positivos. Caso o indivíduo erre as 3 primeiras questões, o critério utilizado será F. Quando não acontece os casos anteriores ou o número de itens administrados já é maior do que 10, volta-se a utilizar o método de melhor desempenho, o KLP.

O artigo citado contribui para o trabalho aqui proposto ao utilizar uma metodologia em que a solução pode ser vista como uma política restrita, em que para cada estado decide-se qual critério de seleção usar e a ação é a regra escolhida.

Outro artigo que envolve ideias que contribuem para a proposta é o trabalho em [El-Alfy \(2011\)](#), que explora uma aplicação de técnicas de RL para testes de domínio sequencial (Sequential Mastery Testing - SMT), como é o caso dos CATs. O objetivo é classificar cada pessoa examinada, usando o número mínimo de itens de teste, como “mestre” ou “não mestre”. Usando RL, um agente inteligente aprende autonomamente a partir das interações para administrar testes de tamanho variável mais informativos e eficazes.

O trabalho compara a aplicação de RL com dois métodos que foram usados e estudados por pesquisadores para encontrar regras de decisão sequencial ótimas: o teste de razão de probabilidade sequencial (SPRT) e teoria da decisão Bayesiana sequencial. O SPRT é relevante em situações onde os itens são disponibilizados sequencialmente. Os itens de teste são selecionados aleatoriamente de um banco de itens calibrados, as respostas do examinando aos itens de teste são tratadas como uma sequência de ensaios de Bernoulli independentes e identicamente distribuídos e o SPRT é usado como um critério de encerramento do teste. A teoria de decisão Bayesiana com TRI foi usada para classificar examinandos em duas categorias. Essa teoria quando aplicada para testes de tamanho variável permite que os custos associados à administração de itens adicionais sejam levados em consideração.

No sistema de teste de domínio sequencial, sob a metodologia de RL, um agente de tomada de decisão interage com os examinandos. O tomador de decisão consiste em dois módulos. O primeiro módulo (SE) estima um estado de crença com base nas respostas do examinando a itens previamente administrados e conhecimento prévio. O estado para cada examinando é estimado sequencialmente a partir da estimativa anterior e da resposta para o item administrado recentemente. Portanto, a cada passo de tempo n , o agente observa a

resposta atual do examinando z_n para o item administrado q_n e então estimará um novo estado θ_n como função de z_n , q_n e θ_{n-1} . O segundo módulo (π) decide se deve parar o teste e declarar a categoria do examinando ou continuar o teste. Sempre que a decisão for continuar o teste, este módulo selecionará um item apropriado, $q_n + 1$, do conjunto de itens no banco de itens (B) que não foram apresentados anteriormente. O conjunto de todas as regras de decisão é a política π que pode ser predeterminada, aproximada e atualizada online, ou completamente aprendida online. O sistema descrito é mostrado na Figura 4.

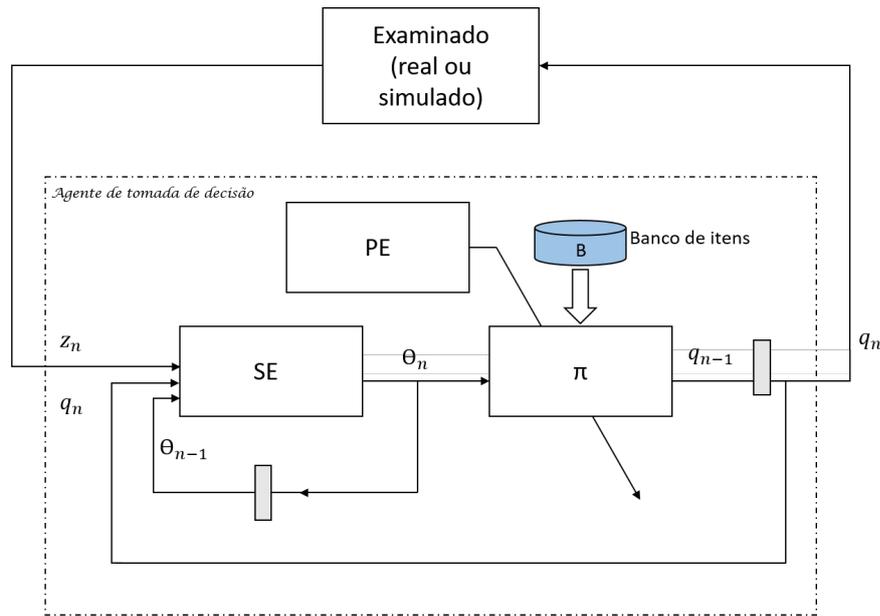


Figura 4 – Sistema de teste de domínio sequencial com RL.

Fonte – Adaptado de [El-Alfy \(2011\)](#)

A sequência de interação com os examinandos se divide em subsequências. Cada subsequência representa a interação com um examinando, onde o sistema começa em algum estado inicial e termina em um estado final. O agente começa a interagir com o examinando j no tempo t onde o estado é s_j^t . Uma ação a_j^t faz com que o estado mude para s_j^{t+1} e receba uma recompensa r_j^t . Então, o agente realiza uma ação a_j^{t+1} e a interação continua até o estado final s_j^{t+T} . O estado do sistema é definido pelo número de respostas corretas após k itens e é denotado por s_k . Definiu-se $d = 2k - s_k$ como a diferença entre o número de respostas corretas e o número de respostas incorretas. O estado do sistema é completamente definido por d se estiver em um estado não terminal d e mais um item é apresentado ao examinando, então, com base em sua resposta, o próximo estado será $d + 1$ ou $d - 1$ dependendo se a resposta está correta ou incorreta, respectivamente. Em cada estado, foram definidas três ações: “continuar”, declarar “mestre”, declarar “não

mestre”. Se o tomador de decisão decidir continuar, haverá um pagamento igual ao custo de administração de um item. Se o tomador de decisão decidir parar e declarar o nível de domínio, um custo de classificação é dado com base no fato da classificação estar correta ou incorreta. Este sistema satisfaz a propriedade de Markov, pois o próximo estado é determinado pelo estado atual e pela ação executada.

Dado o estado do sistema, a distribuição preditiva da resposta do próximo item é definida pelas probabilidades de transição. A fim de determinar a estratégia de teste ideal usando RL, o tomador de decisão define uma função de utilidade que determina o valor esperado de ação i no estado s sob a política π . A melhor ação a ser realizada no estado s é aquela que tem o valor mínimo de ação. O tomador de decisão aprende sequencialmente a política durante a interação com os examinandos.

Usando dados simulados e diversas medidas de desempenho, o estudo avaliou o comportamento da abordagem com RL em comparação com SPRT e teoria da decisão Bayesiana sequencial. Verificou-se que os testes baseados em RL tem melhores taxas de falsos positivos e falsos negativos do que a abordagem Bayesiana. Embora RL tenha um comprimento médio de teste um pouco maior do que as abordagens SPRT e Bayesiana, ele apresenta ser melhor considerando a precisão e a correlação entre as categorias verdadeiras e preditas.

Os resultados obtidos com a abordagem de RL são promissores quando comparados a outras técnicas. Isso motiva uma investigação mais aprofundada do problema sob várias condições, usando MDP e RL.

O estudo em [Nurakhmetov \(2019\)](#) também utilizada uma abordagem de RL para CAT. Neste caso, o algoritmo de seleção sequencial de itens utilizado incorpora o RL em tarefas de classificação. Além disso, é apresentada uma investigação de como CATs podem ser considerados como um processo de decisão sequencial. O estudo explora a possibilidade de resolver problemas, como controle de exposição e balanceamento de conteúdo, que podem ocorrer em CATs, e simultaneamente tenta otimizar a seleção de itens.

O método proposto realiza uma seleção sequencial de itens que aprende quais itens são mais informativos, escolhendo o próximo item em função dos itens já administrados e da crença interna do classificador. Um estudo de simulação mostrou a eficiência do método para testes que requerem uma decisão de classificação confiável com o mínimo de itens possível. No teste de classificação, a seleção de itens é necessária como opção chave para bons resultados de classificação: filtrar itens menos informativos, ao mesmo tempo em que

se presta atenção à exposição e ao controle de conteúdo. Reformulando para o aprendizado de máquina clássico: a seleção de itens é um problema de otimização combinatória que tenta identificar esses itens, o que minimizará o erro de generalização, com diferentes restrições. Em particular, a seleção de itens pode ser vista como um processo que tenta reduzir a quantidade de dados redundantes.

Transformar a classificação em um processo de decisão sequencial resulta na seleção e classificação de itens como um processo adaptativo: decidir qual item selecionar em seguida depende dos itens previamente selecionados e do comportamento do classificador sobre eles. Isso será alcançado usando um classificador treinado como um ambiente para um agente de RL, que aprende qual item selecionar a seguir, recebendo a recompensa na classificação correta do padrão de resposta de entrada parcialmente descoberto.

No primeiro momento, formula-se o CAT como um POMDP, tornando o problema sequencial e, portanto, acessível aos algoritmos de Aprendizado por Reforço. Um requisito do CAT como POMDP foi processar a sequência, uma entrada por vez no modo online, em vez de classificar a sequência inteira de uma vez. Além disso, como saída, um rótulo de classe foi adicionado a cada entrada. Portanto, o classificador sequencial requer algum tipo de memória. Redes Neurais Recorrentes (RNN) são conhecidas por terem memória implícita que pode armazenar informações sobre entradas vistas no passado. Lidar com um POMDP implica precisar extrair uma observação do classificador que resume o passado em uma crença estacionária.

O Estado s no passo de tempo t compreende o padrão de resposta de entrada atual x , o classificador sequencial e o histórico K^* do item anterior h_{t-1} , de modo que $s_t = (x, K^*, h_{t-1})$. A função de recompensa retorna a recompensa r_t no passo de tempo t para a transição do estado s_t para s_{t+1} com ação a_t .

Para minimizar o número de itens (duração do teste) necessários para tomar uma decisão de classificação confiável, é feito um treinamento do agente RL para selecionar itens que levam a uma classificação rápida e precisa. O controle de exposição é mantido usando diferentes recompensas para a seleção de ações, que leva em consideração o histórico de seleção de itens para cada participante do teste. A distribuição de recompensa proposta penaliza o agente durante o processo de treinamento se ele escolher um item que já tenha sido administrado para determinado participante do teste. Além disso, penaliza o agente se o agente selecionar um item cujo limite da taxa de exposição seja excedido.

Por meio de estudos de simulação, o método para seleção adaptativa de itens em CATs mostrou ser eficiente para testes que requerem uma decisão de classificação confiável com o mínimo de itens possível. Além disso, ele pode resolver problemas como superexposição e subexposição com restrições de distribuição de conteúdo, por meio da alocação de recompensa.

Os artigos em [El-Alfy \(2011\)](#) e [Nurakhmetov \(2019\)](#) apresentam abordagens não míopes, porém diferem do trabalho aqui proposto ao tratarem de problemas de classificação.

4 Avaliando critérios de parada

O método de avaliação das estratégias de CAT foi aprimorado. Simulou-se muitos bancos de itens entre diferentes classes de distribuições de itens. Além disso, em vez de considerar um pequeno conjunto de limites previamente definidos para os critérios de parada, foram escolhidos os limites dos critérios de parada após os experimentos.

4.1 Dados sintéticos

Neste trabalho foi considerado o modelo TRI ML3 com os seguintes parâmetros para cada item i : a_i discriminação, b_i dificuldade, c_i probabilidade de uma resposta correta aleatória e $D = 1,702$ fator de escala.

Considerou-se nove classes de distribuição item-banco. Em todos eles os parâmetros de discriminação são extraídos de uma distribuição log-normal $a_i \sim \text{log-normal}(0, 0.35)$ e os parâmetros de probabilidade de uma resposta correta aleatória são extraídos de uma distribuição beta, ou seja, $c_i \sim \text{beta}(1, 4)$. Os parâmetros de dificuldade são extraídos de nove distribuições diferentes.

Seguindo [Babcock e Weiss \(2009\)](#) considerou-se três classes de distribuição para o parâmetro de dificuldade: uniforme, normal e *peak*. A classe uniforme extrai parâmetros de dificuldade de uma distribuição uniforme, ou seja, $b_i \sim \text{uniforme}(-3, 3)$. A classe normal extrai parâmetros de dificuldade de uma distribuição normal padrão, ou seja, $b_i \sim \text{norm}(0, 1)$. A classe de *peak* é uma distribuição mista que com 0,5 de probabilidade extrai itens de uma distribuição normal padrão e com 0,5 de probabilidade extrai itens de uma distribuição uniforme em $(-3, 3)$, conforme proposto no [Babcock e Weiss \(2009\)](#). A distribuição de *peak* simula bancos de itens mais próximos dos reais, onde a maioria dos itens está em torno de uma característica desejada esperada, mas também apresenta itens extremos.

Seguindo [Morris et al. \(2020\)](#), foi aplicado para cada uma das três distribuições – uniforme, normal e *peak* – três níveis de deslocamentos: -1 , 0 e 1 . Observe que, ao considerar -1 como o nível de deslocamento, aumenta-se a probabilidade de ocorrer apenas acertos para alguns alunos de traço alto, já tem-se um número substancial de questões fáceis.

A Figura 5 mostra o intervalo de confiança da informação para três dessas nove distribuições com 100 itens. Observe que, as informações para uma determinada característica θ podem variar substancialmente na distribuição intra-item-banco e na distribuição inter-item-banco.

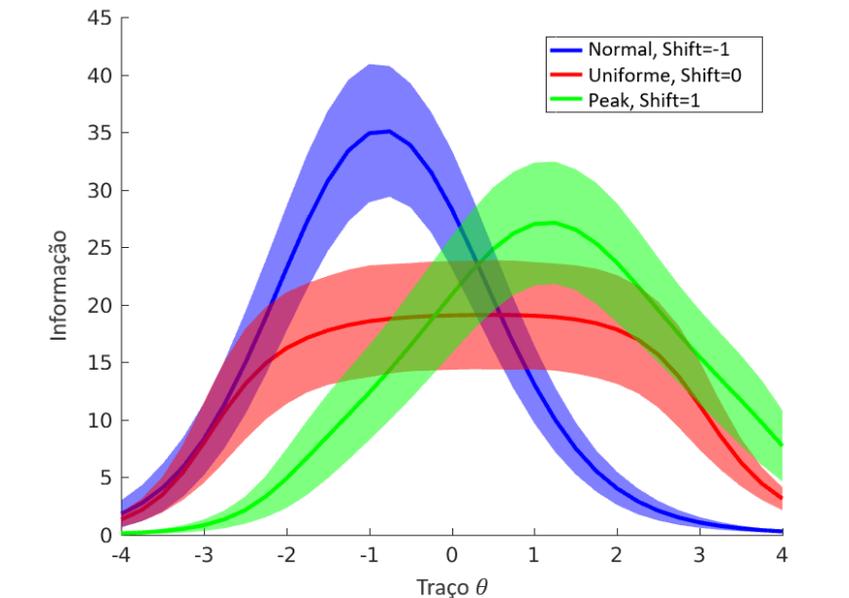


Figura 5 – Informações médias de bancos de 100 itens condicionadas ao traço θ e intervalo de confiança com nível de confiança $\alpha = 0, 1$.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

4.1.1 Simulação CAT

Avaliou-se métodos CAT condicionados a uma das nove classes de banco de itens B . Para cada classe de banco de itens foram feitos:

1. repita 500 vezes:
 - a) retira um banco de 100 itens de B e repete 50 vezes
 - i. fixa um traço θ de uma distribuição normal padrão
 - ii. simula um CAT com duração fixa de 50 questões
 - iii. registra estatísticas relevantes em cada rodada de questões

Experimentou-se os métodos de estimação ML e EAP. Para o método Bayesiano, a distribuição *a priori* considerada para $\Pr(\theta)$ é a distribuição normal padrão. Para o método verossimilhança, a estimativa inicial do traço é 0; caso o aluno erre (acerte) todas

as questões, seu traço latente é diminuído (aumentado) em 0,25 até um mínimo (máximo) de -2 (2). Essa correção é feita porque quando o aluno erra ou acerta todas as questões o θ estimado em ML pode ir para $-\infty$ ou $+\infty$, respectivamente. Após pelo menos um erro e um acerto, o traço é obtido pela máxima verossimilhança.

Em cada rodada n do CAT, registra-se:

- traço estimado $\hat{\theta}_n$;
- erro quadrado $(\hat{\theta}_n - \theta)^2$;
- Erro Padrão $SE(\hat{\theta}_n)$;
- Variância *a posteriori* $VAP(\hat{\theta}_n)$;
- Máxima Informação $MI(\hat{\theta}_n)$;
- Mudança de Theta $CT(\hat{\theta}_n)$; e
- Variância de Variância *a posteriori* $VVAP(\hat{\theta}_n)$.

4.1.2 Equilíbrio entre precisão e eficiência

Normalmente, é fácil obter uma curva de equilíbrio entre precisão e eficiência quando o critério de parada é o Comprimento Fixo (por exemplo, Figuras 6 e 7). Para o critério de parada de comprimento variável, geralmente, um pequeno conjunto de parâmetros de limite é escolhido de antemão.

Uma curva de equilíbrio de eficiência e precisão foi obtida para cada critério de parada escolhendo um conjunto apropriado de níveis de limite. Para cada rodada em cada simulação CAT, registrou-se estatísticas para cada critério de parada. Então, para cada rodada, obteve-se a mediana dessas estatísticas e considerou-se essas medianas como níveis de limiar. Os resultados mostram que a escolha desses níveis de limite dessa maneira permite abranger a curva de equilíbrio ao longo de quase todas as regiões do comprimento do CAT.

Juntamente com o critério de parada, considerou-se duas restrições adicionais: (i) exames não podem parar antes da décima primeira questão; e (ii) exames não podem ter mais de 50 questões.

4.1.3 Resultados

Os resultados obtidos com a metodologia proposta são mostrados a seguir.

Desempenho dos critérios de seleção

Para definir qual critério de seleção de itens usar, comparou-se todos os critérios definidos na Subseção 2.2.1 e definiu-se o critério de parada como FL, com no máximo 30 itens, e observou-se o desempenho de cada critério de seleção. Todas as nove classes do banco de itens foram testadas, aqui discute-se os resultados de uma delas: *peak* com deslocamento de +1.

As Figuras 6 e 7 mostram os resultados para cada critério comparando o comprimento médio do CAT e o RMSE médio, usando os métodos EAP e ML, respectivamente, para estimar a capacidade e *peak* de distribuição para o parâmetro de dificuldade considerando +1 como nível de deslocamento. As Figuras mostram o RMSE absoluto e o RMSE relativo, quando comparados ao método FI.

Utilizando o método EAP, observa-se que os critérios FI, KLP e MEPV são competitivos em relação aos demais. Enquanto MLWI e KL apresentam maior RMSE médio no início do teste; ambos fazem uso da probabilidade para ponderar a média sobre o espaço do traço.

Utilizando o método ML, os critérios KL, KLP e MEPV são mais competitivos que os demais. Neste caso, FI tem um RMSE médio mais alto. O critério MLWI, como no caso anterior, apresenta maior RMSE médio na maioria dos casos.

Todos os critérios de seleção analisados melhoram sua precisão quanto maior o número de questões. Embora o MEPV apresente um bom desempenho, é o mais custoso, pois sempre escolhe o próximo item calculando a variância *a posteriori* considerando todos os itens administrados até então. O FI é menos custoso porque a escolha do item é baseada na estimativa atual de θ .

Portanto, o método FI foi escolhido como critério de seleção de itens para comparar os critérios de parada por ser menos custoso computacionalmente. Embora o método FI não apresente a melhor precisão, a diferença entre os métodos é pequena e acredita-se que não interfira nos resultados seguintes.

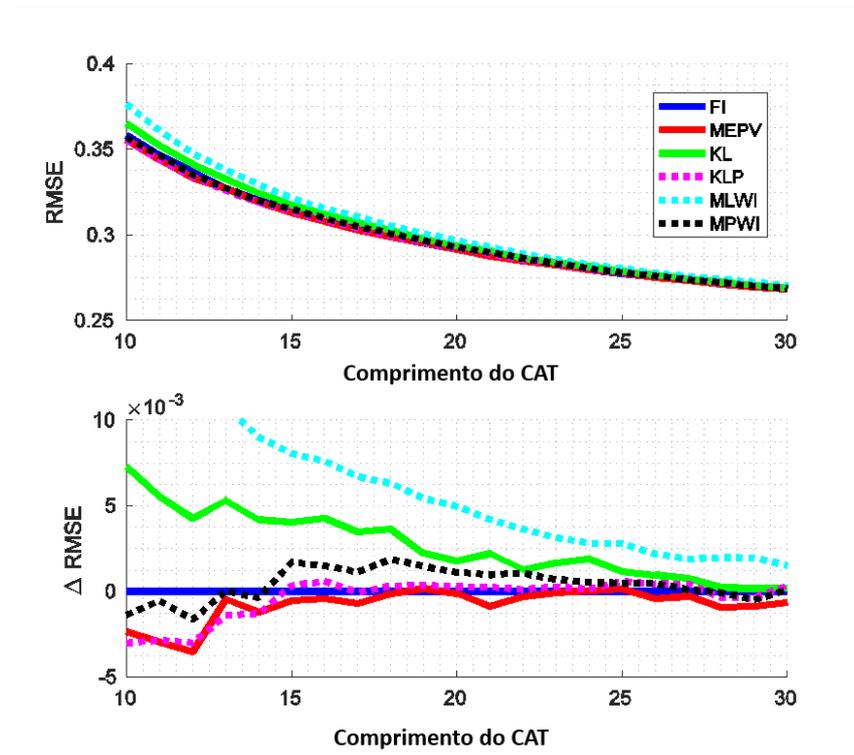


Figura 6 – Comparação dos critérios de seleção de itens utilizando FL como critério de parada, EAP para estimativa e *peak* de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

Desempenho dos critérios de parada

Para comparar o desempenho dos critérios de parada, definindo FI como critério de seleção de itens, o RMSE médio e o comprimento médio do CAT foram calculados usando todos os bancos de itens simulados. O CAT tinha que fazer um mínimo de 10 itens e um máximo de 50, ou seja, se o critério de parada não finalizar o teste até 50 itens, ele é interrompido. Todas as distribuições e deslocamentos nas distribuições definidas para b foram testados.

• Equilíbrio entre precisão e eficiência

Testou-se todas as nove classes do banco de itens, aqui discute-se os resultados de uma delas: *peak* com deslocamento de +1.

As Figuras 8 e 9 mostram os resultados para cada critério de parada comparando o comprimento do CAT e o RMSE médio, usando os métodos EAP e ML, respectivamente, para estimar o traço e *peak* de distribuição para o parâmetro de dificuldade considerando

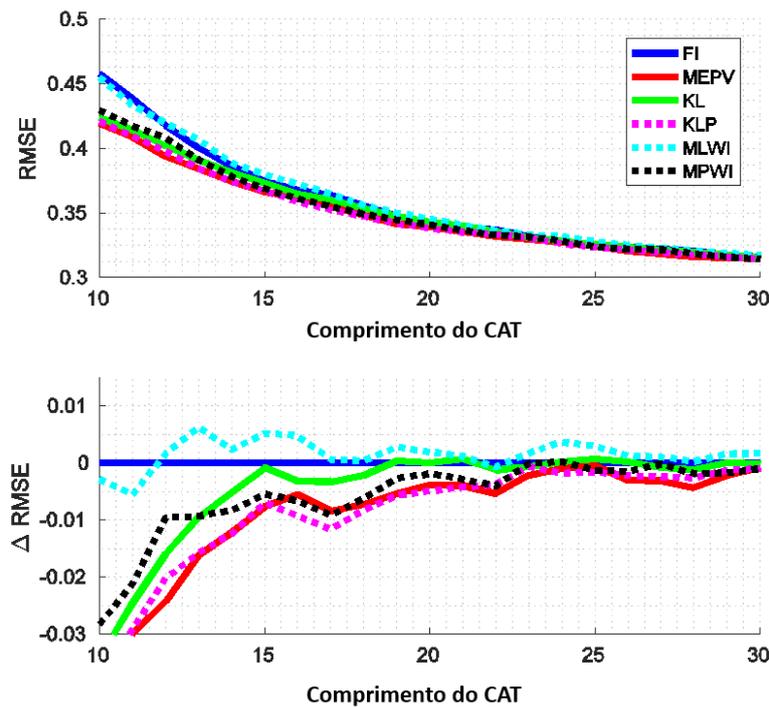


Figura 7 – Comparação dos critérios de seleção de itens utilizando FL como critério de parada, ML para estimativa e *peak* de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

+1 como nível de deslocamento. As Figuras mostram o RMSE absoluto e o RMSE relativo, quando comparados ao método FL.

• Desvio padrão do comprimento CAT

As Figuras 10 e 11 mostram o desvio padrão do comprimento CAT para cada critério de parada usando estimativas EAP e ML, respectivamente.

Como esperado, o desvio padrão para o critério FL é 0. Os critérios com menor variação são MI e CT; ambos avaliam quanta informação resta no banco de itens, o primeiro avaliando as informações, o último avaliando a melhoria na estimativa de características. Métodos baseados em estimativa de variância apresentam o maior desvio padrão; na pior das hipóteses, SE, um CAT com comprimento médio de 30, pode apresentar desvio padrão de 13 itens.

• Comprimento do CAT vs traço

Embora o método FL apresente um equilíbrio competitivo entre precisão e eficiência e nenhuma variância, outro critério de parada pode ser vantajoso se o examinador quiser

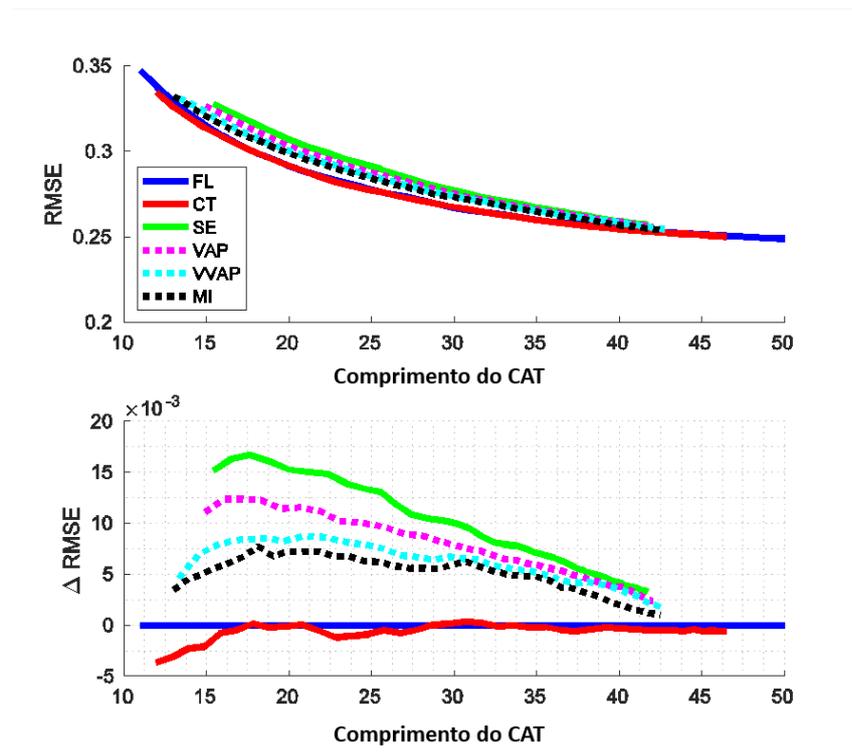


Figura 8 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens, EAP para estimativa e *peak* de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

diferenciar examinandos. Por exemplo, o examinador pode desejar que os examinandos com a característica mais baixa gastem menos tempo respondendo ao teste.

As Figuras 12 e 13 mostram a diferença no comprimento do CAT em relação ao comprimento médio do CAT dividindo o espaço do traço em 5 faixas com a mesma quantidade de traços após seguir a distribuição do desvio padrão. O primeiro gráfico refere-se às características mais baixas e o último gráfico refere-se às características mais altas.

Considerou-se a distribuição *peak* com nível de deslocamento igual a -1 (Figura 12) e +1 (Figura 13). No primeiro caso, há menos informações para examinandos no 5° intervalo, enquanto no segundo caso, há menos informações para examinandos no 1° intervalo (Figura 5).

Crítérios de parada baseados em informações, como MI e CT, utilizam menos itens para examinandos com poucas informações no banco de itens; por outro lado, critérios de parada baseados em variância, como SE, VAP e VVAP, utilizam menos itens para

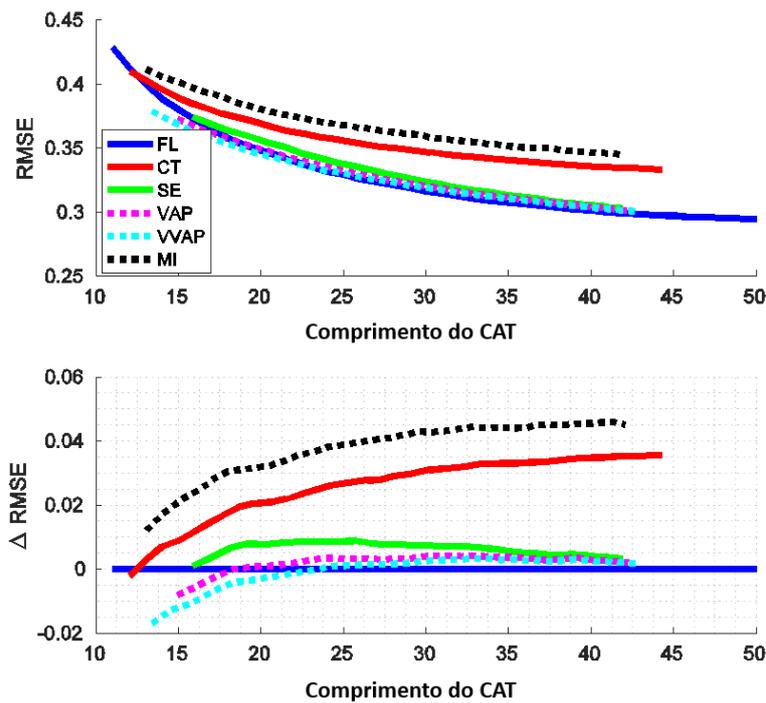


Figura 9 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens, ML para estimativa e *peak* de distribuição para parâmetro de dificuldade considerando +1 como nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

examinandos com grande informação no banco de itens. Figuras 12 e 13 mostram esse comportamento oposto.

• Combinando Critérios de Parada

Como cada método de critério de parada apresenta características diferentes, pode-se considerar a combinação de dois ou mais métodos para apresentar um melhor desempenho. Babcock e Weiss (2009) e Morris *et al.* (2020) consideram a combinação *ad hoc*. Aqui, considera-se a otimização baseada na curva de equilíbrio proposta.

Considerando novamente a partição de examinandos em cinco intervalos. As Figuras 14 e 15 mostram o RMSE de cada método de critério de parada em cada intervalo para estimativas EAP e ML, respectivamente. Escolheu-se a classe de banco de itens com base na distribuição normal e nível de deslocamento igual a -1 ; esta foi a classe onde observou-se maior diferença por meio dos intervalos, de forma que nenhum método dominou os demais.

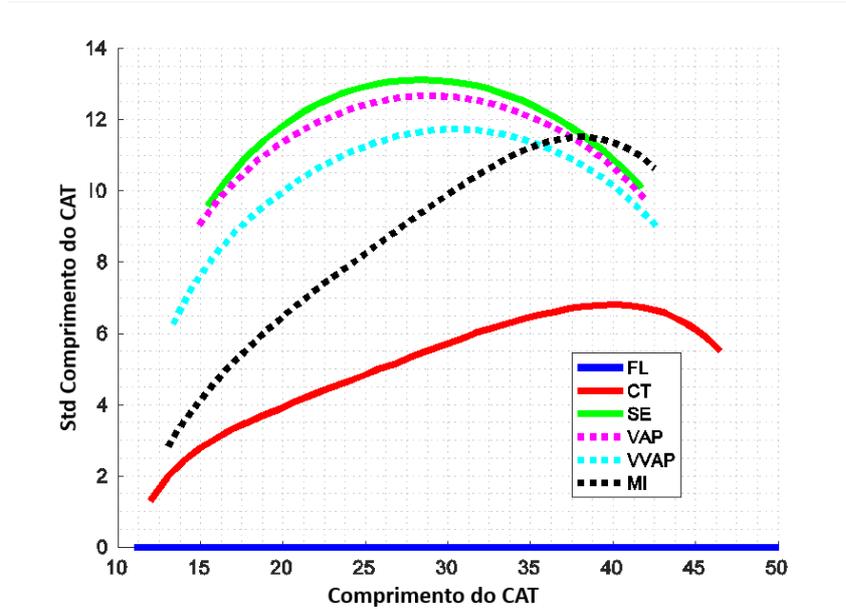


Figura 10 – Desvio padrão de comprimento CAT para cada critério de parada usando estimativa EAP e *peak* de distribuição para parâmetro de dificuldade considerando +1 como o nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

Considerou-se duas combinações de critérios de parada: oráculo e estimado. No caso de oráculo é um caso irreal, quando o examinador conhece a faixa de onde vem o examinado e pode escolher o melhor critério de parada para cada examinado; portanto, a melhoria é garantida. No caso de estimado, o examinador estima o traço do examinado com base na resposta aos itens aplicados e escolhe o critério de parada *online*, alterando quando necessário.

As Figuras 14 e 15 mostram o resultado médio de ambos os critérios de parada mistos, usando EAP e ML na estimativa, com base na habilidade real (pontos vermelhos) e habilidade estimada (pontos verdes).

Os resultados mostram que quando diferencia-se os indivíduos, ou seja, considera-se os níveis de características separadamente, é possível melhorar o RMSE combinando os critérios de parada. A melhoria é mais destacada quando se considera a estimativa de ML.

Embora o método baseado em oráculo apresente alguma melhoria, a melhoria não é grande, mesmo quando considera-se a melhor classe de banco de itens a apresentar tal melhoria. Por outro lado, o método baseado em estimativas nem mesmo apresenta melhoria com a estimativa EAP, o que pode mostrar que a estimativa é muito imprecisa para ser usada como guia para critério de parada condicional ou critério de seleção.

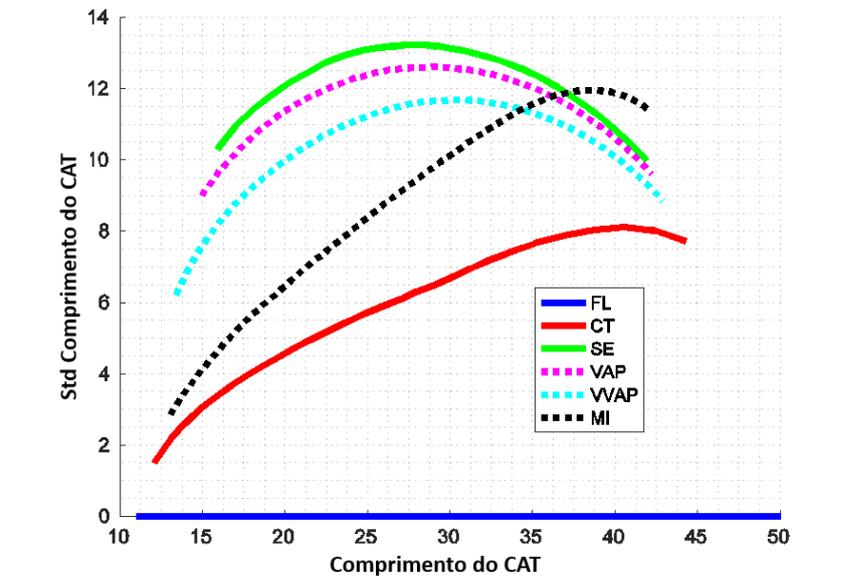


Figura 11 – Desvio padrão de comprimento CAT para cada critério de parada usando estimativa de ML e *peak* de distribuição para parâmetro de dificuldade considerando +1 como o nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

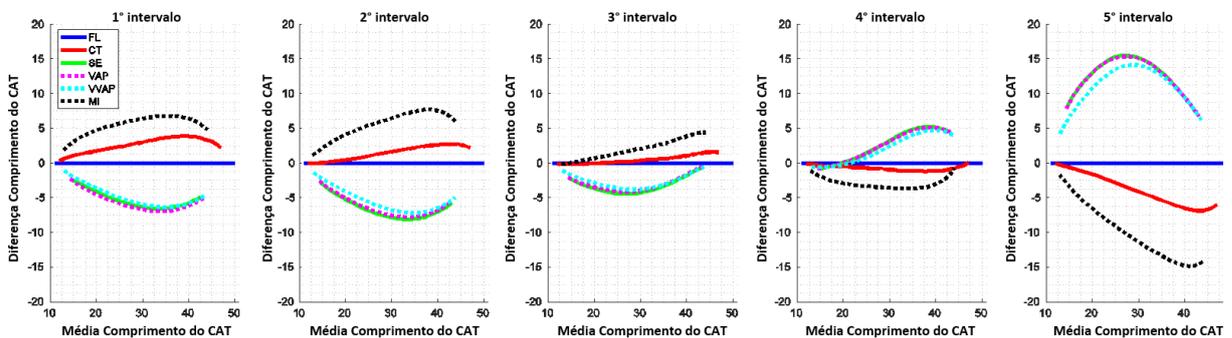


Figura 12 – Diferença no comprimento do CAT em relação ao comprimento médio do CAT, dividindo a distribuição do traço em 5 faixas usando estimativa de EAP e *peak* de distribuição para parâmetro de dificuldade considerando -1 como o nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

4.2 Dados reais

Foram utilizados os dados da prova do ENEM de 2012, os quais são públicos e foram retirados do portal da transparência (TEIXEIRA, 2016). O modelo de dados é composto por um conjunto de respostas dicotômicas de alunos selecionados de forma aleatória. Cada

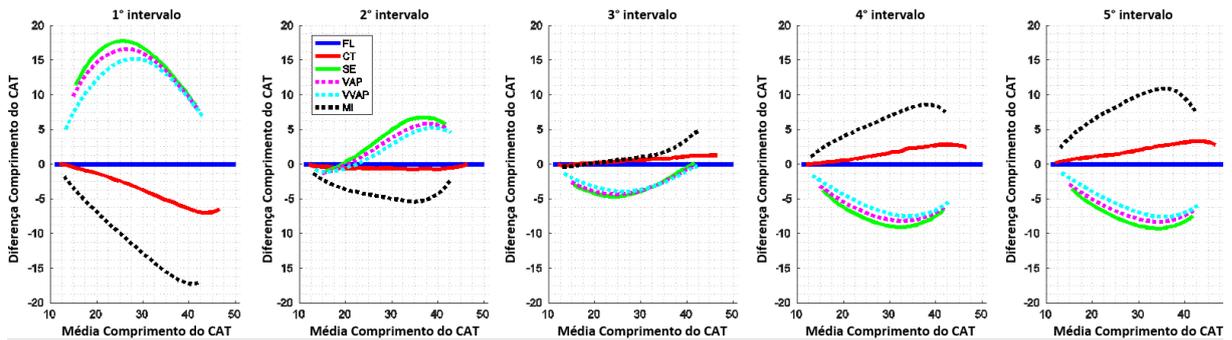


Figura 13 – Diferença no comprimento do CAT em relação ao comprimento médio do CAT, dividindo a distribuição do traço em 5 faixas usando estimativa de EAP e *peak* de distribuição para parâmetro de dificuldade considerando +1 como o nível de deslocamento.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

resposta u , pode conter apenas dois possíveis valores: 1 para “acertou” e 0 para “errou”, ou seja, $u = 0, 1$.

A amostra foi composta por 10.000 alunos da prova Rosa de Matemática e suas tecnologias. Os valores dos parâmetros a , b e c das 45 questões foram estimados com o modelo ML3 e 20 pontos de quadratura por meio do pacote *irtoys* (PARTCHEV; PARTCHEV; SUGGESTS, 2017) do *software* R Core Team (2022).

Neste caso, experimentou-se o método de estimativa ML. A estimativa inicial do traço é 0; caso o aluno erre (acerte) todas as questões, sua característica é diminuída (aumentada) em 0,25 até um mínimo (máximo) de -3 (3); após pelo menos um erro e um acerto, o traço é obtido pela máxima verossimilhança.

O θ real foi obtido com as estimativas ML considerando à prova completa de 45 questões.

Resultados

Para comparar o desempenho dos critérios de parada, definindo FI como critério de seleção de itens, o RMSE médio e o comprimento médio do CAT foram calculados. O CAT tinha que fazer um mínimo de 10 itens e um máximo de 45, ou seja, se o critério de parada não finalizar o teste até 45 itens, ele é interrompido.

- **Equilíbrio entre precisão e eficiência**

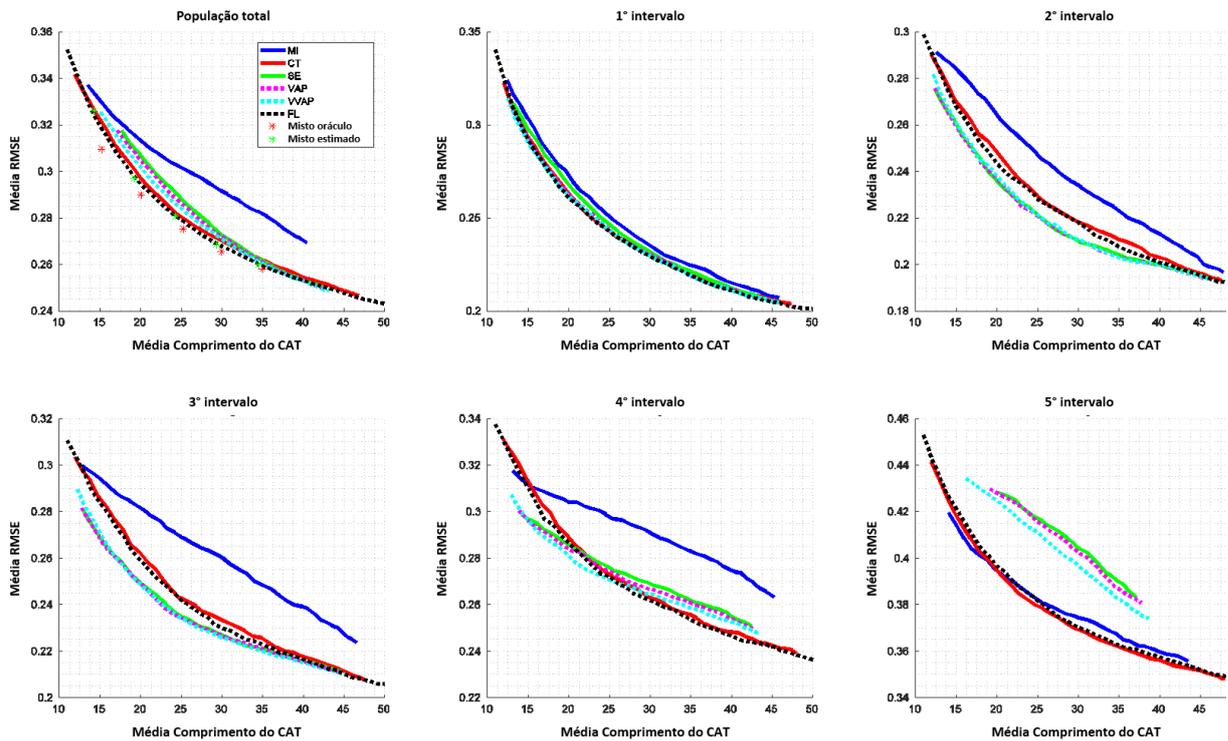


Figura 14 – Média de todos os critérios de parada misturados usando EAP na estimativa de capacidade.

A Figura 16 mostra os resultados para cada critério de parada comparando o comprimento do CAT e o RMSE médio, usando o método ML para estimar o traço.

• **Desvio padrão do comprimento CAT**

A Figura 17 mostra o desvio padrão do comprimento CAT para cada critério de parada usando estimativa ML.

Como no caso dos dados simulados, o desvio padrão para o critério FL é 0. Os critérios com menor variação são VVAP e CT. Tanto aqui como no caso de dados simulados SE apresenta o maior desvio padrão.

• **Comprimento do CAT vs traço**

A Figura 18 mostra a diferença no comprimento do CAT em relação ao comprimento médio do CAT dividindo o espaço do traço em 5 faixas com a mesma quantidade de traços após seguir a distribuição do desvio padrão. Como no caso dos dados simulados, o primeiro gráfico refere-se às características mais baixas e o último gráfico refere-se às características mais altas.

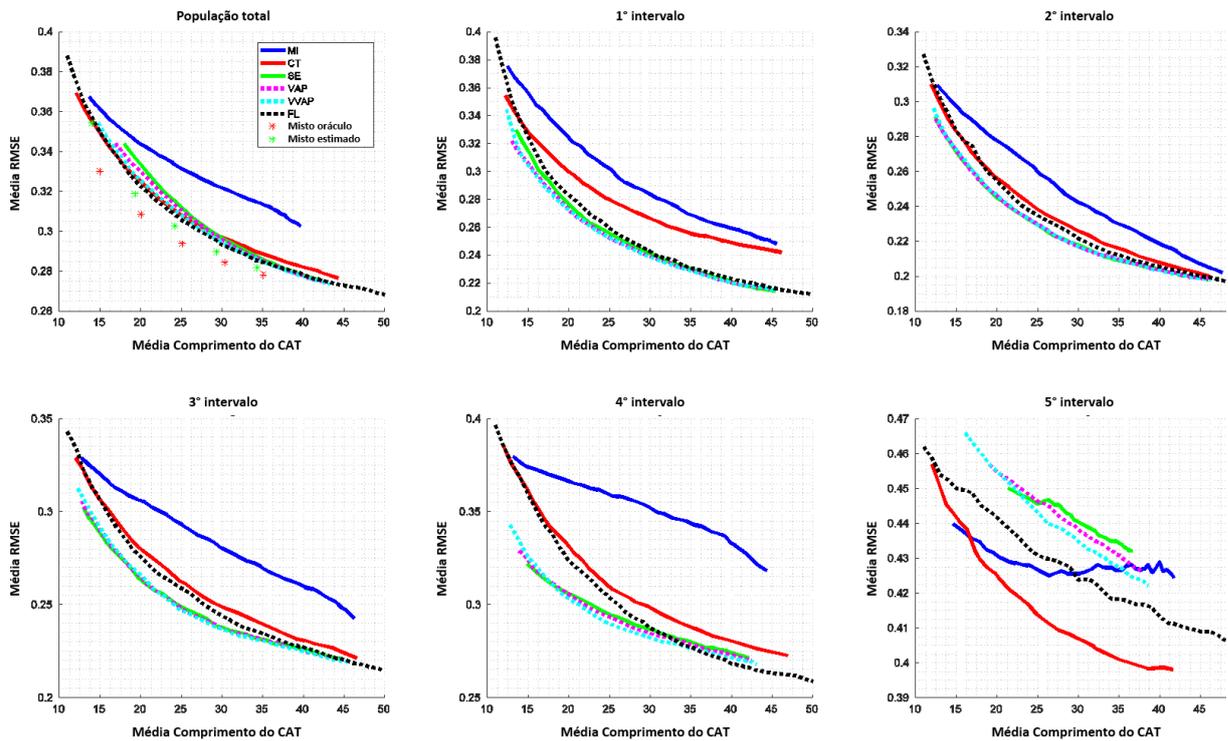


Figura 15 – Média de todos os critérios de parada misturados usando ML na estimativa de capacidade.

Considerando novamente a partição de examinandos em cinco intervalos. As Figuras 19 e 20 mostram o RMSE e o Desvio padrão de comprimento CAT, respectivamente, de cada método de critério de parada em cada intervalo para estimativas ML.

Crítérios de parada baseados em informações, como MI e CT, utilizam menos itens para examinandos com poucas informações no banco de itens; por outro lado, critérios de parada baseados em variância, como SE e VAP, utilizam menos itens para examinandos com grande informação no banco de itens. O VVAP mostrou ter um desempenho bom para todos os intervalos e seu desvio padrão é o segundo menor, perdendo apenas para o FL.

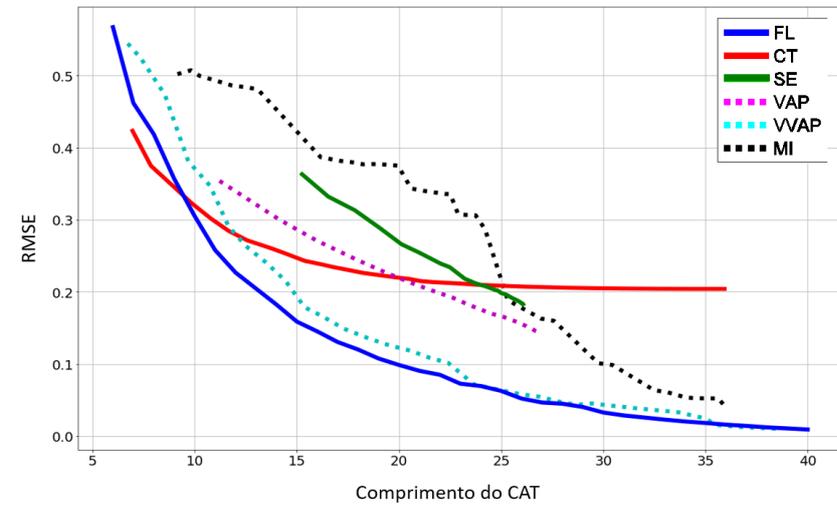


Figura 16 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens e ML para estimativa.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

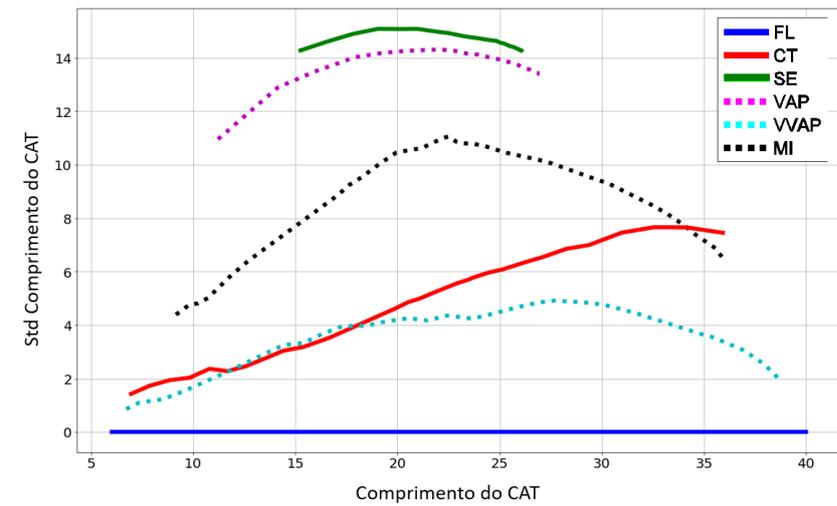


Figura 17 – Desvio padrão de comprimento CAT para cada critério de parada usando estimativa de ML.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

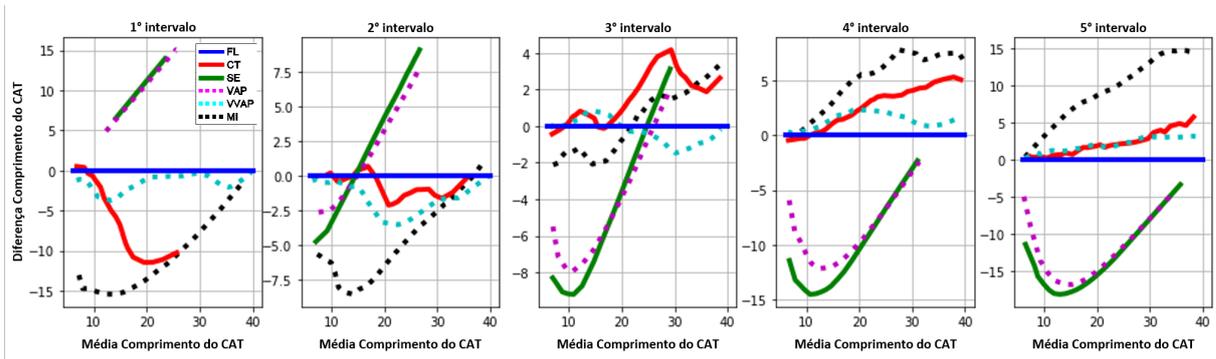


Figura 18 – Diferença no comprimento do CAT em relação ao comprimento médio do CAT, dividindo a distribuição do traço em 5 faixas usando estimativa de ML.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

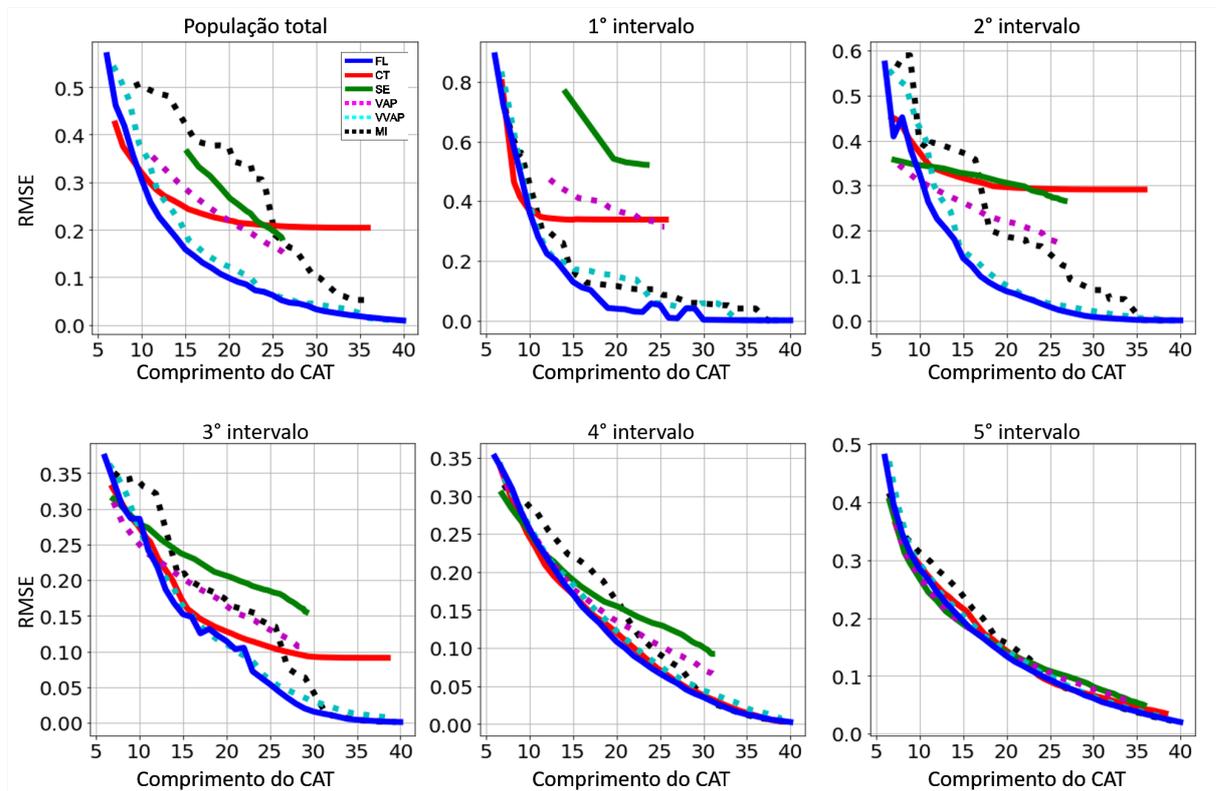


Figura 19 – Comparação dos critérios de parada utilizando FI como critério de seleção de itens, dividindo a distribuição do traço em 5 faixas usando estimativa de ML.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

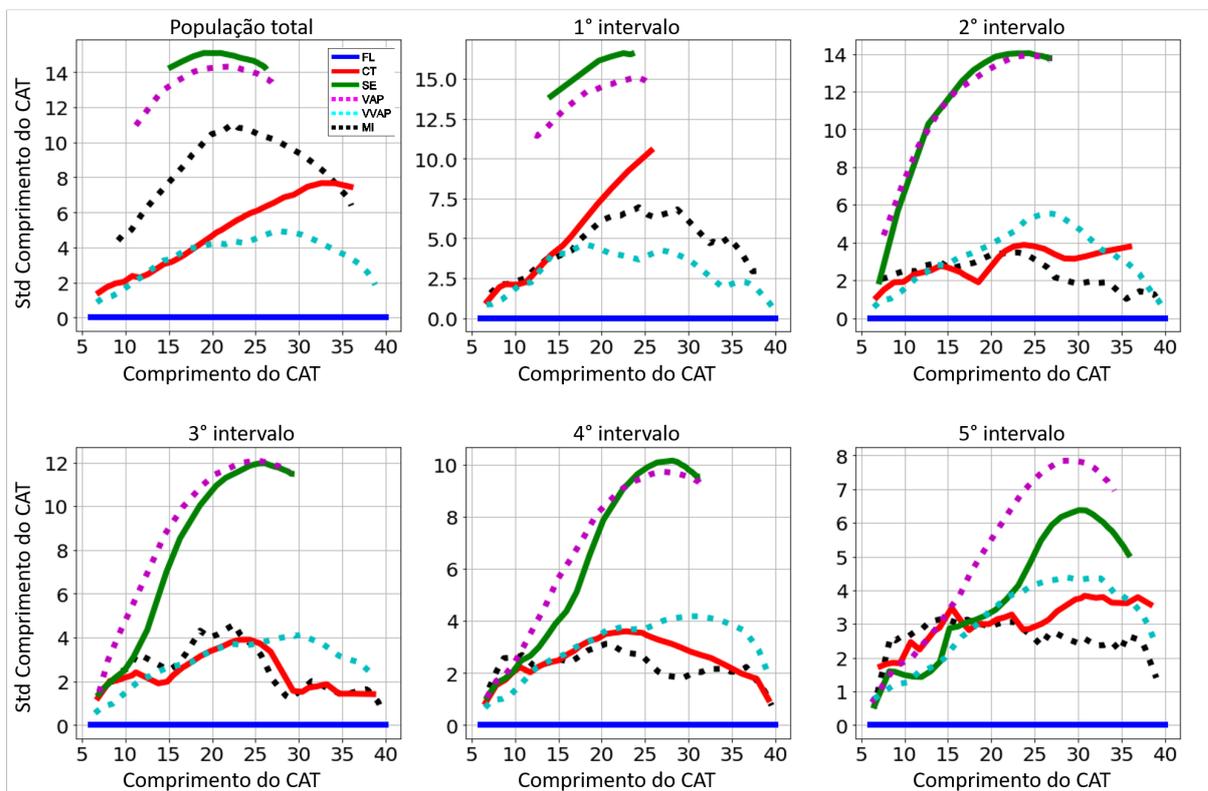


Figura 20 – Desvio padrão de comprimento CAT para cada critério de parada, dividindo a distribuição do traço em 5 faixas usando estimativa de ML.

Fonte – Patricia Gilavert e Valdinei Freire, 2021

5 CAT como um problema de otimização

Dado um banco de itens Q , primeiro modela-se o problema CAT como um processo:

1. Um examinando θ é extraído de uma distribuição $f(\theta)$. Embora a distribuição $f(\theta)$ possa ser conhecida pelo examinador, o processo não revela o examinando θ .
2. Enquanto o examinador não decide parar o teste, em qualquer estágio $t = 1, 2, \dots$
 - a) o examinador escolhe uma questão $q_t \in Q$ e submete ao examinado¹.
 - b) Ao responder a questão, o examinando gera um resultado x_t .
3. O examinador estima o traço latente θ por meio de uma estimativa $\hat{\theta}$ ².

O resultado do processo CAT é um histórico aleatório $h = (\theta, q_1, x_1, q_2, x_2, \dots, q_N, x_N, \hat{\theta})$. Um examinador ótimo é obtido a partir do seguinte problema de otimização:

$$\begin{aligned} \min \quad & E_{\theta}[U(h)] \\ \text{s.t.} \quad & Z(f(\theta), h) \end{aligned}$$

U é uma função objetivo. Por exemplo, se o examinador quiser minimizar o MSE, então, $U(h) = (\theta - \hat{\theta})^2$. Se o examinador quiser minimizar a duração do teste, então, $U(h) = N$. A função objetivo também pode combinar os dois objetivos.

Z é uma função de restrição. Se o examinador quiser restringir o item à exposição máxima α , então, $Z(\cdot) = \{h | E[\sum_{t=1}^N 1_{q_t=q}] \leq \alpha \forall q \in Q\}$. Se o examinador quiser evitar a repetição de questões, então, $Z(\cdot) = \{h | \sum_{t=1}^N \sum_{t'=t+1}^N 1_{q_t=q_{t'}} = 0 \forall q \in Q\}$. Observe que no primeiro exemplo, as restrições estão na expectativa sobre a variável aleatória h , enquanto no segundo exemplo as restrições estão nas amostras da variável aleatória.

5.1 CAT como POMDP

No processo de um CAT, inicialmente um examinando é extraído de uma distribuição de probabilidade $f(\theta)$. Então, diante de qualquer questão, o examinador observa a resposta para aquela questão, que depende da questão em si e do traço latente θ do examinando. No entanto, o examinador nunca observa diretamente o traço latente θ . Um CAT pode ser visto como um POMDP por:

¹ No CAT multiestágio, o examinador deve escolher um conjunto de questões.

² Em uma classificação CAT, o examinador pode estimar diretamente uma categoria ao invés de um traço latente θ .

1. Considerando o espaço de estados como a tupla (θ, q_t) , ou seja, o traço latente e a última questão.
2. Considerando o espaço de ação como o conjunto de questões no banco de itens, mais uma ação de rescisão.
3. Considerando transições determinísticas. Primeiro, θ nunca muda e a questão q_t muda de acordo com as escolhas do examinador.
4. Considerando a observação do processo como resultado da resposta do examinando à última questão, ou seja, o modelo de resposta CAT.

A Figura 21 mostra um esquema conectando CAT e POMDP. Lá, o estado de crença é obtido a partir da distribuição *a posteriori* baseada no método Bayesiano. I é o estado de crença, enquanto τ é o modelo para gerar o estado de crença *a posteriori*.

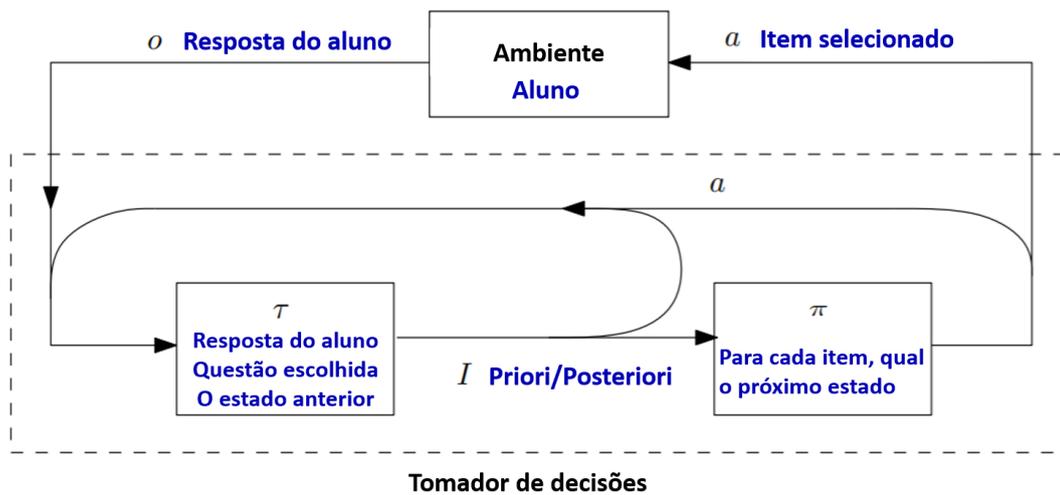


Figura 21 – CAT enquadrado em um POMDP com estado de crença.

Fonte – Patricia Gilavert, 2022

• Minimizar MSE com N de tamanho fixo sem questões repetidas

Ainda não formulou-se a função de custo. Essencialmente, depende do compromisso entre precisão e eficiência. Neste cenário, deseja-se obter o menor MSE (função de utilidade U) com no máximo N questões (função de restrição Z).

Para isso, deve-se aumentar o espaço de observação com todas as questões submetidas ao examinando e respectivos resultados. Então, um custo terminal é pago com base na variância da distribuição *a posteriori*. O espaço de ação depende da observação, ou seja,

$a_t \in \mathcal{A}(o_t)$ e se uma questão já foi administrada, tal questão não está incluída em $\mathcal{A}(o_t)$. Após N questões, a única questão disponível em $\mathcal{A}(o_t)$ é a ação de rescisão.

- **Controle de exposição do item: toda ação é exposta igualmente**

Considere a formulação LP1 na Seção 3.1, se todas as questões devem ser expostas igualmente, a seguinte restrição deve ser considerada para cada par $a, a' \in \mathcal{A}$:

$$\sum_{s \in \mathcal{S}} x_{s,a} = \sum_{s \in \mathcal{S}} x_{s,a'}.$$

$x_{s,a}$ indica a quantidade esperada de ocorrência do par (s, a) . A soma $\sum_{s \in \mathcal{S}} x_{s,a}$ é a ocorrência total esperada para a ação a , portanto, uma questão a .

- **CAT de vários estágios: K questões por estágio**

Neste caso, uma ação é um subconjunto do banco de itens Q com K questões, ou seja, $\mathcal{A} = \{A \in 2^Q \text{ e } |A| = K\}$, onde 2^Q é o conjunto de potências de Q .

Observe que embora as formulações aqui apresentadas possam não ser práticas para serem resolvidas de forma otimizada, elas são descritas sob a mesma estrutura, MDPs e POMDPs. Na próxima seção é apresentado um resultado prático da formulação teórica.

5.2 Aproximando um CAT por um MDP

Nesta seção mostra-se que o resultado teórico apresentado na seção anterior pode elucidar resultados na literatura CAT. Considerou-se um problema CAT clássico: comprimento fixo com um banco de itens dicotômico. Aproximou-se este CAT por um MDP, que pode ser resolvido de forma ótima, e mostrou-se que o critério de seleção tradicional, como Informação de Fisher, pode estar próximo da solução ótima.

Modelo de Resposta: Teoria de Resposta ao Item

Considerou-se o ML3 (BIRNBAUM, 1968). Um banco de 45 itens calibrados de um exame nacional de matemática foi usado no experimento (SPENASSATO *et al.*, 2016).

5.2.1 Método Bayesiano Normalizado e Discretizado

Como dito anteriormente, uma maneira de recuperar a propriedade de Markov em um processo é considerar o estado aumentado ou estado de crença. A distribuição *a posteriori* seguindo o método Bayesiano é exatamente o estado de crença do CAT POMDP. No entanto, o espaço de estados de tal distribuição é contínuo \mathbb{R}_1 . Um MDP finito a partir de tais estados de crença foi construído.

Primeiro, normalizou-se cada distribuição *a posteriori* por uma distribuição normal com média e variância equivalentes. Segundo, discretizou-se tal par de valores; considerou-se 100 valores médios (entre -4 e 4) e 1.000 valores de variância (entre 0,001 e 1). Finalmente, as transições são definidas entre os estados de crença discretizados normalizados.

Normalmente, um CAT não repete um item no mesmo exame. Portanto, o espaço de estados deve ser aumentado com as questões aplicadas. Neste caso, o número de estado é exponencial no número de questões. Para reduzir o número de estados, permitiu-se que o CAT repita as questões; neste caso, o espaço de estados não precisa ser aumentado. As respostas às questões são probabilísticas; portanto, o usuário pode responder de forma diferente para a mesma questão.

Considerou-se o problema de otimização CAT de minimizar MSE com N de comprimento fixo com questões repetidas. Nesse caso, uma política ótima pode ser facilmente encontrada. Na próxima seção, compara-se com a abordagem tradicional da literatura CAT.

5.2.2 Resultados

Considerou-se dois métodos de seleção de itens da literatura CAT: Informação de Fisher (FI) (LORD, 1980; SARI; RABORN, 2018) e Minimizando a Variação Posterior Esperada (MEPV) (MORRIS *et al.*, 2020). O primeiro, FI, é bem conhecido na literatura e também é o método mais barato. A segunda, MEPV, é uma função de pontuação dispendiosa, que considera os próximos estados de crença potenciais dadas todas as questões possíveis no banco de itens; é ótimo em nosso MDP discretizado normalizado quando apenas uma questão é permitida. Definiu-se políticas com base em cada função de pontuação e avaliou-se sob a estrutura do MDP; por causa da estrutura do MDP, a função de valor V^π para cada função de pontuação pode ser calculada exatamente.

As funções de pontuação são míopes, na verdade, elas nem levam em conta o horizonte de comprimento fixo N . Elas são claramente subótimas, no entanto, está longe de ser claro o quão longe da política ideal elas estão. Utilizou-se um algoritmo MDP para obter uma política ótima para cada comprimento entre 1 e 45 questões. Todas as outras políticas, comparou-se com esse padrão-ouro.

Figure 22 mostra os resultados. Traçou-se a diferença entre o MSE raiz (RMSE) para cada política e o padrão-ouro. Além das políticas FI e MEPV, também traçou-se o comportamento das políticas ótimas para comprimento fixo 15, 30 e 45 e o comportamento quando o MDP decide onde parar. Observe que as políticas ótimas podem estar longe do padrão-ouro durante as primeiras questões, mas no final (15, 30 ou 45 questões) o padrão-ouro é alcançado. Como as políticas baseadas em funções de pontuação escolhem questões que dão retorno imediato, elas estão sempre próximas do padrão-ouro, mas nunca são ótimas. Apesar da função de pontuação não ser ótima, pode-se observar que a maior diferença não é superior a 0,01. No ultimo caso, em que o MDP escolhe um item e decide se para, observa-se um compromisso melhor entre erro e quantidade de questões.

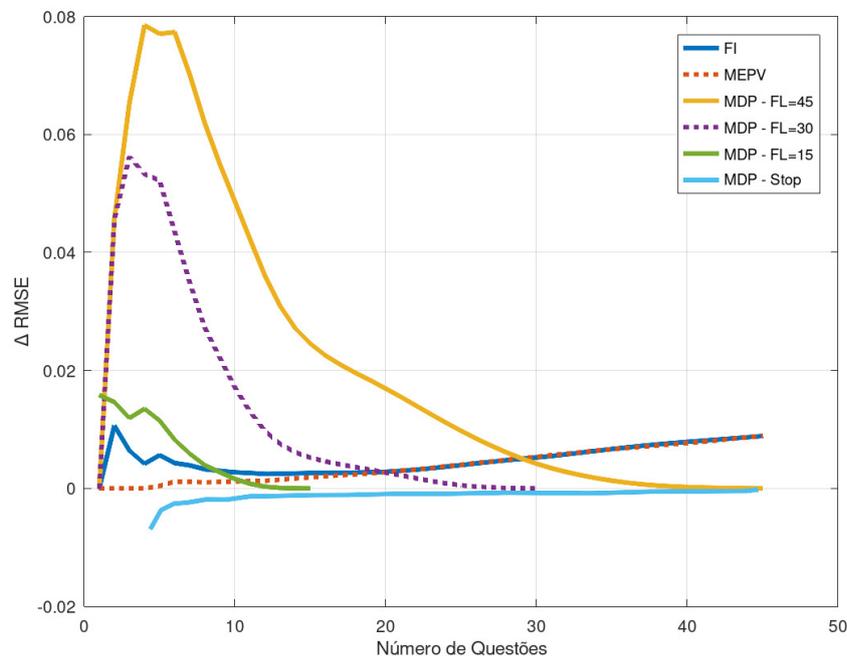


Figura 22 – Comparação de políticas baseadas em funções de pontuação míope e políticas ótimas de MDP.

6 Conclusão e considerações finais

O critério de parada proposto, VVAP, apresenta desempenho semelhante à maioria dos demais critérios para dados simulados. Para dados reais, o critério tem desempenho melhor e por vezes consegue acompanhar FL, porém é pior por apresentar maior desvio padrão no número de questões. Esta é também uma vantagem do critério Comprimento Fixo para todos os outros critérios considerados.

Embora muitos trabalhos utilizem critérios de parada mistos, observou-se que não parece melhorar o RMSE médio quando se considera a população total.

Concluí-se a favor do critério FL, desde que possa ser ajustado ao banco de itens em questão. O FL mostra uma curva de equilíbrio de precisão e eficiência competitiva em todos os cenários, apresentando variação zero na duração do teste.

Os métodos de definição de limites apresentados foram importantes para comparar de forma justa todos os critérios em cada banco de itens.

A limitação da pesquisa atual é fixar apenas o ML3 da TRI para o cálculo da probabilidade de pontuação correta. Trabalhos futuros podem ser desenvolvidos usando outros modelos de TRI e outros bancos de itens reais.

Elaborou-se formalismos CAT como MDPs e mostrou-se em um experimento que apesar de ótimo, o ganho com MDPs pode não compensar, pois soluções subótimas devem ser consideradas em cenários reais. No entanto, a formulação como MDPs permite formular CAT como problema de otimização e descrever muitas formulações de CAT sob a mesma estrutura.

Acredita-se que uma estrutura e experimentos como o aqui mostrado podem elucidar os limites de uma grande quantidade de métodos na literatura de CAT, principalmente no que se refere a CAT sob restrições. Recentemente, os autores mostraram que o critério de parada de comprimento fixo tem grandes vantagens em relação a outros critérios de parada, com esta estrutura pode-se investigar como ele se compara à otimalidade sob diferentes avaliações (GILAVERT; FREIRE, 2021). Por exemplo, a estrutura MDP permite definir a otimização sensível ao risco (MINAMI; SILVA, 2012); Os MDPs sensíveis ao risco permitiriam que um CAT ponderasse os piores cenários em relação à duração do teste ou ao MSE, de modo que o CAT fosse justo para todos os examinandos.

Este trabalho resultou na publicações de dois artigos *Comprehensive Empirical Analysis of Stop Criteria in Computerized Adaptive Testing* (GILAVERT; FREIRE, 2021) e *Computerized Adaptive Testing: a unified approach under Markov Decision Process* (GILAVERT; FREIRE, 2022).

Os códigos implementados e bases de dados deste trabalho estão disponíveis no link [GitHub-CAT-MDP](#).

Referências

- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. *ABE, São Paulo*, 2000. Citado na página 23.
- BABCOCK, B.; WEISS, D. J. Termination criteria in computerized adaptive tests: Variable-length cats are not biased. In: *D. J. Weiss (Ed.)*. USA: Conference on Computerized Adaptive Testing, 2009. v. 1, p. 1–18. Citado 5 vezes nas páginas 20, 27, 28, 41 e 48.
- BAKER, F. B.; KIM, S.-H. *Item response theory: Parameter estimation techniques*. Boca Raton, Florida, EUA: CRC Press, 2004. Citado na página 22.
- BERTSEKAS, D. P.; TSITSIKLIS, J. N. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, INFORMS, Institute for Operations Research and the Management Sciences (INFORMS), Linthicum, Maryland, USA, v. 16, n. 3, p. 580–595, ago. 1991. ISSN 0364-765X. Citado na página 31.
- BIRNBAUM, A. L. Some latent trait models and their use in inferring an examinee’s ability. *Statistical theories of mental test scores*, Addison-Wesley, 1968. Citado 2 vezes nas páginas 22 e 59.
- CHANG, H.-H.; YING, Z. A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 3, p. 213–229, 1996. Citado na página 26.
- EL-ALFY, E.-S. M. A reinforcement learning approach for sequential mastery testing. p. 295–301, 2011. Citado 4 vezes nas páginas 18, 36, 37 e 40.
- GILAVERT, P.; FREIRE, V. Comprehensive empirical analysis of stop criteria in computerized adaptive testing. In: *Proceedings of the 13th International Conference on Computer Supported Education*. Online streaming: Science and Technology Publications, 2021. Citado 2 vezes nas páginas 62 e 63.
- GILAVERT, P.; FREIRE, V. Computerized adaptive testing: A unified approach under markov decision process. In: *International Conference on Computational Science and Its Applications (CSEDU)*. Malaga, Espanha: Springer, 2022. p. 591–602. Citado na página 63.
- GOSAVI, A. Reinforcement learning: A tutorial survey and recent advances. *INFORMS Journal on Computing*, INFORMS, v. 21, n. 2, p. 178–192, 2009. Citado na página 31.
- HAMBLETON, R. K.; SWAMINATHAN, H. *Item response theory: Principles and applications*. New York, USA: Springer Science & Business Media, 2013. Citado na página 17.
- HOERGER, M.; KURNIAWATI, H. *An On-Line POMDP Solver for Continuous Observation Spaces*. 2020. Citado 2 vezes nas páginas 19 e 34.
- JATOBÁ, V. M. G.; FARIAS, J. S.; FREIRE, V.; RUELA, A. S.; DELGADO, K. V. Alicat: a customized approach to item selection process in computerized adaptive testing. *Journal of the Brazilian Computer Society*, Springer, v. 26, p. 1–13, 2020. Citado na página 35.

- KAEHLING, L. P.; LITTMAN, M. L.; CASSANDRA, A. R. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, Elsevier, v. 101, n. 1-2, p. 99–134, 1998. Citado na página 30.
- KREITZBERG, C. B.; STOCKING, M. L.; SWANSON, L. Computerized adaptive testing: Principles and directions. *Computers & Education*, Elsevier, v. 2, n. 4, p. 319–329, 1978. Citado na página 17.
- LINDEN, W. J. van der. Bayesian item selection criteria for adaptive testing. *Psychometrika*, Springer, v. 63, n. 2, p. 201–216, 1998. Citado na página 26.
- LINDEN, W. J. van der; GLAS, C. A. *Computerized Adaptive Testing: Theory and Practice*. Boston, MA: Springer Science & Business Media, 2000. Citado na página 24.
- LORD, F. M. *Applications of item response theory to practical testing problems*. New York, USA: Routledge, 1980. Citado 3 vezes nas páginas 19, 26 e 60.
- MAGIS, D.; YAN, D.; DAVIER, A. A. von. *Computerized Adaptive and Multistage Testing with R: Using Packages CatR and MstR*. 1st. ed. USA: Springer Publishing Company, Incorporated, 2017. ISBN 3319692178. Citado na página 29.
- MAUSAM; KOLOBOV, A. Planning with markov decision processes: An ai perspective. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, v. 6, n. 1, 2012. Citado 2 vezes nas páginas 18 e 32.
- MINAMI, R.; SILVA, V. F. d. Shortest stochastic path with risk sensitive evaluation. In: *11th Mexican International Conference on Artificial Intelligence (MICAI 2012)*. Berlin: Springer, 2012. (Lecture Notes in Artificial Intelligence, v. 7629), p. 370–381. Citado na página 62.
- MORRIS, S. B.; BASS, M.; HOWARD, E.; NEAPOLITAN, R. E. Stopping rules for computer adaptive testing when item banks have nonuniform information. *International journal of testing*, Taylor & Francis, v. 20, n. 2, p. 146–168, 2020. Citado 5 vezes nas páginas 20, 26, 41, 48 e 60.
- NURAKHMETOV, D. Reinforcement learning applied to adaptive classification testing. In: *Theoretical and Practical Advances in Computer-based Educational Measurement*. Cham: Springer, 2019. p. 325–336. Citado 3 vezes nas páginas 18, 38 e 40.
- PARTCHEV, I.; PARTCHEV, M. I.; SUGGESTS, M. Package ‘irtoys’. *A collection of functions related to item response theory (IRT)*, 2017. Citado na página 51.
- PELLEGRINI, J.; WAINER, J. Processos de decisão de markov: um tutorial. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 133–179, 2007. Citado na página 30.
- PUTERMAN, M. L. *Markov decision processes: discrete stochastic dynamic programming*. New Jersey: John Wiley & Sons, 2014. Citado na página 33.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>. Citado na página 51.
- RASCH, G. Probabilistic models for some intelligence and achievement tests. University of Chicago Press. Copenhagen, Denmark: Danish Institute for Educational Research, 1960. Citado na página 22.

- SARI, H. I.; RABORN, A. What information works best?: A comparison of routing methods. *Applied psychological measurement*, SAGE Publications Sage CA: Los Angeles, CA, v. 42, n. 6, p. 499–515, 2018. Citado 2 vezes nas páginas 26 e 60.
- SEGALL, D. O. Computerized adaptive testing. *Encyclopedia of social measurement*, Academic Press New York, NY, v. 1, p. 429–438, 2005. Citado na página 17.
- SPENASSATO, D.; BORNIA, A.; TEZZA, R. Computerized adaptive testing: A review of research and technical characteristics. *IEEE Latin America Transactions*, v. 13, n. 12, p. 3890–3898, 2015. Citado 2 vezes nas páginas 17 e 24.
- SPENASSATO, D.; TRIERWEILLER, A. C.; ANDRADE, D. F. de; BORNIA, A. C. Testes adaptativos computadorizados aplicados em avaliações educacionais. *Revista Brasileira de Informática na Educação*, v. 24, n. 02, p. 1, 2016. Citado na página 59.
- STAFFORD, R. E.; RUNYON, C. R.; CASABIANCA, J. M.; DODD, B. G. Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior research methods*, Springer, v. 51, n. 3, p. 1305–1320, 2019. Citado na página 28.
- SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998. ISBN 0-262-19398-1. Citado na página 18.
- TEIXEIRA, I. N. *Microdados do ENEM. Brasília: Inep, 2016*. 2016. <http://portal.inep.gov.br/basica-levantamentos-acessar>. Acesso em: 29 mar. 2017. Citado na página 50.
- TREVIZAN, F.; TEICHTIL-KÖNIGSBUCH, F.; THIÉBAUX, S. Efficient solutions for stochastic shortest path problems with dead ends. Australian, 2017. Citado na página 32.
- VEERKAMP, W. J.; BERGER, M. P. Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, Sage Publications Sage CA: Los Angeles, CA, v. 22, n. 2, p. 203–226, 1997. Citado na página 26.
- WAINER, H.; DORANS, N. J.; FLAUGHER, R.; GREEN, B. F.; MISLEVY, R. J. *Computerized adaptive testing: A primer*. New York: Routledge, 2000. Citado na página 17.
- WANG, C.; CHANG, H.-H.; HUEBNER, A. Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement*, Wiley Online Library, v. 48, n. 3, p. 255–273, 2011. Citado na página 17.