Provas Ótimas

Roberto Marques Bekman

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do grau
de
Mestre em Estatística

Área de Concentração: Estatística
Orientador: **Prof. Dr. Sergio Wechsler**

São Paulo - junho - 2005

Provas ótimas

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Roberto Marques Bekman e aprovada pela comissão julgadora.

São Paulo, junho de 2005.

Banca examinadora:

- . Prof. Dr. Sérgio Wechsler (orientador) (IME-USP)
- . Prof. Dr. Luis Gustavo Esteves (IME-USP)
- . Profa. Dra. Verônica Andréa Gonzalez-Lopez (UNICAMP)

*À minha esposa Margarete, aos meus
filhos Karen e Felipe.*

*Agradecimentos ao prof. Sérgio, à
Fundação Carlos Chagas pelo apoio
oferecido e a ANBID e SESI pelos
dados disponibilizados para o estudo.*

SUMÁRIO

Provas e Exames são rotineiramente utilizados para se inferir a habilidade de um examinando ou mesmo se inferir se um paciente está ou não doente. No texto iremos apresentar o Experimento de Máxima Variância de Bayes como o melhor ensaio a fim de revelar nossa incerteza sobre a habilidade de um examinando. Os ensaios abordarão a escolha de um único item (questão de uma prova); a escolha de uma prova completa (com um número fixado de itens) e provas adaptadas ao examinando (onde busca-se a próxima questão com base nas informações acumuladas até o momento). O texto abordará desde o desenvolvimento teórico da escolha do(s) melhor(es) experimento(s), no caso a(s) melhor(es) questão(ões) de uma prova, até a análise de casos reais, incluindo técnicas de simulação necessárias ante a dificuldade matemática da obtenção da solução analiticamente exata.

Palavras-chave: Inferência Bayesiana, Informação, Experimento de Máxima Variância de Bayes, Teoria da Decisão, Decisão Ótima de Bayes, Simulação da Densidade Posterior, Provas Ótimas, Provas Adaptadas por Computador.

ÍNDICE

1) Introdução	1
2) O Conceito de Informação	4
3) O Experimento de Máxima Variância de Bayes	8
4) O Experimento Ótimo à Luz da Teoria da Decisão	11
5) O Problema da Escolha dos k Melhores Experimentos	14
6) O Problema da Escolha Seqüencial dos k Melhores Experimentos	20
7) A Teoria da Resposta ao Item	22
8) Aplicação Prática na Escolha dos Melhores Itens de uma Prova	44
9) Conclusão	63
Bibliografia	64

1) Introdução

A realização de provas e exames fazem parte do cotidiano de praticamente todas as pessoas, na vida moderna, em quase toda a extensão de suas vidas. Provas são realizadas rotineiramente durante o ensino fundamental, ensino médio, passando pelos vestibulares e cursos de nível superior. Não devemos esquecer que a realização de provas também faz parte de uma gama muito maior de atividades, englobando cursos de língua estrangeira, cursos de extensão profissional, exames de habilitação (de condutor de veículos, embarcações, aviação), concursos públicos, exames de certificação profissional, exames de admissão empresarial, etc. Podemos acomodar um número ainda maior de aplicações ao considerarmos exames de análises clínicas ou ainda qualquer outro problema em que, em determinado instante, faz-se necessário escolher um experimento (prova, exame, ensaio) dentre outros experimentos disponíveis a fim de obtermos informação a respeito de um estado da natureza cujo valor verdadeiro não é conhecido.

Nesse contexto, em que a realização de uma prova se faz necessária para aferir as aptidões individuais com os mais diversos objetivos, torna-se desejável desenvolver métodos baratos, eficientes e precisos para se aferir tais aptidões. Nessa direção, a introdução das questões de múltipla escolha constituiu, sem dúvida, um grande passo em termos de objetividade e baixo custo na estimação das habilidades do público-alvo nos assuntos desejados pelo avaliador.

Focado nos testes de múltipla escolha, o presente trabalho irá apresentar um método estatístico para a decisão de quais questões deverão integrar uma prova de modo que os resultados obtidos sejam os melhores possíveis. Essa preocupação se justifica pois o senso comum nos diz que questões (itens da prova) diferentes possuem dificuldades intrínsecas diferentes; some-se ainda que num teste de múltipla escolha existe sempre uma possibilidade de se acertar uma questão por mera sorte (acerto casual); do mesmo modo, um examinando com grande domínio no assunto colocado em teste pode errar uma questão por mero descuido. Assim, em casos extremos, se uma prova contém apenas questões extremamente difíceis, estaremos medindo, em sua prevalência, a distribuição do acerto casual dos examinandos; no outro extremo, caso a prova contenha apenas questões demasiadamente fáceis, estaremos medindo, em sua prevalência, a distribuição do erro por descuido, ou seja, nestes casos extremos não estamos obtendo nenhuma informação relevante acerca da habilidade dos examinandos. Numa prova usual, esse problema normalmente é contornado incluindo-se, de forma empírica, um certo número mínimo de questões consideradas fáceis/médias/difíceis pelo próprio professor que elabora a prova. Deste modo, empírico, o professor direciona a escolha das questões segundo seus conhecimentos com o objetivo de obter informação sobre as habilidades dos examinandos.

Para formalizar os conceitos descritos acima, iremos introduzir a Teoria da Resposta ao Item, (D Andrade (1998), R Klein (2003) e R Valle (2000)), método razoavelmente bem conhecido e difundido no ramo de avaliação e certificação antes de avançar na direção de se escolher as questões ideais de uma prova através da Teoria da Decisão Estatística.

A escolha das questões ideais de uma prova, fundamentada na Teoria da Decisão Estatística, constituirá o principal objetivo desse trabalho e se valerá de um banco de questões previamente testadas e analisadas segundo a Teoria da Resposta ao Item. A partir desse banco de questões e de uma informação a priori sobre o grupo ou examinando a ser avaliado, será desenvolvido um método ótimo para a escolha das questões de uma prova com um número fixo de questões e, a partir das mesmas informações a priori, será desenvolvido, adicionalmente, um método para escolha ótima seqüencial dos itens de uma prova para cada examinando.

O método de escolha dos itens acima citado se valerá dos princípios da Inferência Bayesiana e da Teoria da Decisão: as incertezas sobre as habilidades dos examinandos θ_i , latentes, não mensuráveis diretamente, serão representadas por $f(\theta_i)$, a priori, e por $f(\theta_i | x_i)$, a posteriori, onde i representa um determinado examinando i e x_i são as respostas (acertos e erros) nas diversas questões a ele. Será apresentada uma função de perda associada à escolha do item X_{ij} para a avaliação de θ_i , condizente com a teoria da decisão. Essa função de perda refletirá uma operação sobre a decisão de se escolher X_{ij} para reduzir a incerteza sobre θ_i diante da opção de escolher um outro item j para o examinando i num determinado instante. Será verificado que a função de perda definida reflete a informação sobre θ_i contida no experimento X_{ij} . Desta forma, será verificado que o melhor experimento X_{ij}^* (apresentar o item j^* ao indivíduo i) é aquele que produz a maior variância do estimador de Bayes $V(E[\theta | X_{ij}])$. (M DeGroot (1962)).

Os resultados obtidos serão aplicados na composição de provas otimizadas à luz da Teoria da Decisão Estatística, tanto para provas "Tradicionais", onde um conjunto de itens deverá ser escolhido, em grupo, antes da obtenção de qualquer informação x_{ij} , como também em provas "Adaptativas por Computador" onde os dados $x_i = (x_{i1}, x_{i2}, \dots, x_{ij})$ são conhecidos antes da escolha de $X_{i,j+1}$. Essa metodologia estatística, uma vez aplicada em larga escala, possibilitaria ganhos adicionais em toda a cadeia produtiva:

As provas do tipo "Tradicional", como se sabe, são utilizadas amplamente. Nesse tipo de prova, as questões são selecionadas e diagramadas para então serem reproduzidas em alta tiragem (impressas em papel) e aplicadas para um grande número de pessoas. Nesse segmento, a escolha das questões que comporão a prova com o método (estatístico) apresentado neste trabalho permitirá que as provas contenham menos questões que as provas usuais, garantindo-se uma incerteza equivalente a respeito das habilidades dos examinandos. Isso nos permitirá alcançar ganhos com a economia de papel, tempo de impressão, custo de transporte, menor tempo de aplicação, menor tempo de correção. Ou seja, a escolha das questões de uma prova "Tradicional", valendo-se do método apresentado, implicará na redução dos prazos e nos custos, pois as provas conterão menos questões e, portanto, menos páginas, em média, e menos dados amostrais a serem computados, sem perda de capacidade inferencial.

Já testes do tipo "Testes Adaptativos por Computador" são um segmento a se desenvolver no país, sendo que importantes instituições no exterior já se utilizam do computador para apresentar uma versão *on-line* da prova. Testes Adaptativos por Computador utilizam-se de análises estatísticas *on-line* dos erros e acertos das questões anteriores, respondidas pelo examinando, para subsidiar a decisão de qual será a próxima questão a se apresentar ao examinando que está realizando a prova à frente do computador. Essencialmente, se um examinando acerta uma questão, o bom senso nos diz que seria indicado escolher uma questão mais difícil para apresentar a seguir e vice-versa. Nesse segmento, a escolha das questões pelo método apresentado no presente trabalho indicará, através de análises estatísticas, a melhor questão do banco a se apresentar em cada etapa da prova, ponderando-se as informações acumuladas dos erros e acertos do examinando anteriormente. Esse método permitirá alcançar uma incerteza equivalente a respeito das habilidades dos examinandos com um número ainda menor de questões que na prova "Tradicional" ótima do parágrafo anterior. Como os procedimentos de

aplicação de um "Teste Adaptativo por Computador" e de um teste "Tradicional", são essencialmente diversos entre si, a comparação de prazos e custos imediatos fica comprometida; contudo podemos listar algumas vantagens muito interessantes do "Teste Adaptativo por Computador" ante o "Teste Tradicional": (1) as provas de dois examinandos distintos tendem a ser diferentes entre si, inviabilizando a cola. Apesar das provas de dois examinandos serem potencialmente diferentes entre si, os resultados serão plenamente comparáveis; (2) como as provas tendem a ser diferentes umas das outras de acordo com a habilidade dos examinandos, as provas poderiam ser realizadas em horários e datas diferentes, por agendamento, sem que isso compromettesse a confiabilidade e credibilidade dos resultados obtidos; (3) é possível realizar uma prova com um número menor de questões que na prova "Tradicional" sem perda de capacidade inferencial.

Em resumo, a escolha das questões que constituirão uma prova pode ser tratada estatisticamente auferindo-se, com os métodos propostos neste trabalho, ganhos adicionais em eficiência e/ou flexibilidade em relação aos procedimentos tradicionais. Esses ganhos serão obtidos mediante análises estatísticas, eventualmente *on-line*, que avaliam o próximo item a ser apresentado aos examinandos. Mesmo com provas menores (menos questões), será possível manter ou até mesmo melhorar a precisão na estimação das habilidades dos examinandos que se submeterem aos exames segundo a metodologia apresentada para a montagem das provas. Essa conjectura se torá evidente quando verificarmos, no último capítulo, uma quantidade significativa de questões numa prova "Tradicional" que "revelam" praticamente nenhuma informação sobre a habilidade dos examinandos.

Para acomodar esses temas, o trabalho será desenvolvido em 3 etapas

Primeira Etapa: Teórica

Nessa etapa, que abrange os capítulos 2, 3 e 4, os seguintes tópicos são abordados:

- A abordagem Teórica do Experimento de maior Informação de Bayes
- A abordagem Teórica da Teoria da Decisão Estatística
- Estabelece uma **conexão** entre essas duas abordagens

Segunda Etapa: Teórica

Nessa etapa, que abrange os capítulos 5 e 6, os seguintes tópicos são abordados:

- A escolha dos k Melhores Experimentos
- A escolha sequencial dos k Melhores Experimentos

Observação: esses temas, ainda teóricos, aplicam-se a várias áreas do conhecimento

Terceira Etapa: Teórico-Prática

Nessa etapa, que abrange os capítulos 2, 3 e 4, os seguintes tópicos são abordados:

- Introdução à Teoria da Resposta ao Item
- Apresentação de um método de simulação da distribuição da habilidade de um examinando
- Comparação e validação da simulação com dados de outro programa comercial (BILOG)
- Aplicação da simulação na escolha dos Melhores Itens de uma Prova Objetiva
- Análise de caso entre provas realizadas

2) O Conceito de Informação

Informação, numa visão generalista de D Basu (1975), é o que ela faz por você, muda a sua opinião.

Informação, no âmbito da Estatística, é qualquer dado amostral (resultado de um teste, exame, ensaio) que provoque alguma mudança no conhecimento de algum parâmetro (estado da natureza) de interesse cujo valor verdadeiro não é conhecido. Entenda-se estado da natureza, no nosso caso, como sendo a verdadeira habilidade de um examinando.

Representaremos tal estado da natureza por θ , a habilidade de um examinando. Diante a nossa incerteza a respeito de θ , iremos representá-la por uma densidade de probabilidade, a priori, $f(\theta)$. Essa densidade reflete o julgamento, com base nas informações então disponíveis, sobre o verdadeiro valor de θ . Há um caráter subjetivo nessa abordagem; normalmente esse julgamento, a priori, acima referido, é emitido por um especialista da área, ou ainda, reflete a sua opinião sobre o verdadeiro estado na natureza. Devemos entender a densidade $f(\theta)$ como sua sincera opinião acerca do verdadeiro valor do estado na natureza θ . A realização de um experimento (teste, exame, ensaio) é representada por $X = x$, onde x é um dos possíveis resultados do experimento, no nosso caso, dicotômico. Quando um experimento ainda não foi realizado, representaremos-o por X , uma variável aleatória.

Dessa forma, estamos considerando que o acerto ($X = 1$) e erro ($X = 0$) de um examinando submetido a uma determinada questão está correlacionado com sua habilidade θ , cujo verdadeiro valor não é mensurável diretamente. Em outros termos, as variáveis X (acerto e erro numa determinada questão) e θ (habilidade) são estatisticamente dependentes; onde X é mensurável diretamente e, θ , não. Usaremos X a fim de inferir θ .

Quando avaliamos as respostas de um examinando submetido a um conjunto de experimentos $X_J = (X_{j1}, X_{j2}, \dots, X_{jk})$, futuramente apresentada como uma prova com k questões, estamos considerando que $(X_{j1}, X_{j2}, \dots, X_{jk}, \theta)$ tem uma distribuição conjunta estatisticamente dependente, onde X_J é o conjunto de itens $X_{j1}, X_{j2}, \dots, X_{jk}$ submetido a um examinando. Paralelamente será considerado que, uma vez conhecido o verdadeiro valor de θ , as variáveis aleatórias $X_{j1}, X_{j2}, \dots, X_{jk}$, dado θ , são independentes. Com efeito, uma vez conhecido o verdadeiro valor de θ , a realização de qualquer experimento X torna-se dispensável, pois $P(X = x | \theta)$ é meramente uma probabilidade de acerto ou erro para tal valor conhecido de θ . Em outros termos, uma vez conhecido o verdadeiro valor θ , a realização de qualquer experimento X não traz nenhuma Informação adicional sobre θ .

O primeiro resultado da abordagem estatística para o referido problema é que, uma vez realizado o experimento, devemos atualizar nossa opinião acerca do parâmetro θ mediante a Operação de Bayes:

$$f(\theta | x_j) = \frac{L(\theta | x_j) \cdot f(\theta)}{\int L(\theta | x_j) \cdot f(\theta) \cdot d\theta} \quad (\theta \text{ contínuo}) \quad (i)$$

$$P(\theta | x_j) = \frac{L(\theta | x_j) \cdot P(\theta)}{\sum_{\theta} L(\theta | x_j) \cdot P(\theta)} \quad (\theta \text{ discreto}) \quad (ii)$$

Dessa forma fica claro que a Informação Estatística está contida na verossimilhança $L(\theta | x_j)$ pois é ela que faz mudar nossa opinião $f(\theta)$ a priori para $f(\theta | x_j)$ a posteriori.

Considerando que os experimentos são dicotômicos X_j (erro/acerto, negativo/positivo) podemos considerá-los como experimentos de Bernoulli com parâmetro $h_j(\theta)$: $X_j \sim \text{Bernoulli}(h_j(\theta))$. Usando a propriedade da independência condicional, temos que a verossimilhança de x_j , para um subconjunto J de experimentos, é dada por

$$L(\theta | x_j) = \prod_j [h_j(\theta)]^{x_j} \cdot [1 - h_j(\theta)]^{1-x_j}$$

No modelo logístico de 3 parâmetros da Teoria da Resposta ao Item, abordado futuramente,

$$h_j(\theta) = c_j + \frac{1 - c_j}{1 + e^{-Da_j(\theta - b_j)}}$$

onde a_j , b_j e c_j são os parâmetros que modelam a probabilidade de acerto $h_j(\theta)$ do item j segundo a Teoria da Resposta ao Item, detalhada no capítulo 7.

Nesse caso, o Operador de Bayes para o conjunto J de experimentos é igual a

$$f(\theta | x_j) = \frac{L(\theta | x_j) \cdot f(\theta)}{\int L(\theta | x_j) \cdot f(\theta) \cdot d\theta} \quad (\theta \text{ contínuo}) \quad (iii)$$

$$P(\theta | x_j) = \frac{L(\theta | x_j) \cdot P(\theta)}{\sum_{\theta} L(\theta | x_j) \cdot P(\theta)} \quad (\theta \text{ discreto}) \quad (iv)$$

Notemos que a Informação Estatística permanece contida na verossimilhança, agora $L(\theta | x_j)$.

Definiremos, formalmente, a informação como:

$$I\{X; P, U\} = U(P) - E[U(P_x)] \quad (v)$$

A informação, definida dessa forma, está atrelada à definição de um operador de Incerteza U , operador esse que reflete uma medida de incerteza a respeito do estado da natureza θ . Adicionalmente, X representa um experimento e, P , probabilidade; P , aplicada sobre θ , é a distribuição a priori de θ , representada por $f(\theta)$, enquanto que P , aplicada sobre $\theta | x$, é a distribuição posterior $f(\theta | x)$, obtida mediante Operação de Bayes (i) a (v).

Deste modo, estabelecidas a função operadora de Informação e P , probabilidade, acerca da nossa incerteza sobre o estado da natureza de interesse θ , estamos aptos a medir a Informação do Experimento X mediante sua respectiva verossimilhança. Dessa forma denominaremos por Experimento de Máxima Variância de Bayes, o experimento que maximiza a $I\{X, P; V\} = V(\theta) - V(E[\theta | X])$. Nesse caso, a Variância é o operador de Incerteza U e a Esperança $E[\theta | X = x]$ é o operador sobre a distribuição posterior $\theta | X = x$.

Um aspecto importante a se demonstrar é que a função U , Operador da Incerteza, deve ser côncava. M DeGroot (1962), demonstra que a Informação $I\{X, P; U\} \geq 0$, se e somente se, a Função de Incerteza U é côncava. Esse resultado implica, biunivocamente, que se a Função de Incerteza U é côncava, esperamos uma diminuição na Incerteza sobre o parâmetro (estado da natureza) de interesse com a realização do Experimento X .

Tal demonstração decorre da desigualdade de Jensen para distribuições condicionais:

seja φ uma função convexa. Então $\varphi\{E[Y | Z]\} \leq E\{\varphi(Y) | Z\}$. B James (1996).

Analogamente,

seja φ uma função côncava. Então $E\{\varphi(Y) | Z\} \leq \varphi\{E[Y | Z]\}$.

Dessa forma, tomando Y como sendo a distribuição posterior $\xi(X)$ e Z como ξ , distribuição a priori e U , côncava,

$$E\{U(\xi(X)) | \xi\} \leq U\{E[\xi(X) | \xi]\}, \quad (\text{vi})$$

onde $E[\xi(X) | \xi]$ representa o vetor $(E[\xi_1(X) | \xi], \dots, E[\xi_n(X) | \xi])$, onde n é um dos possíveis valores da priori ξ_i , $i = 1, \dots, n$, mas

$$E[\xi_i(X) | \xi] = \int \xi_i(x) \left[\sum_{j=1}^m \xi_j f_j(x) \right] d\mu = \int \xi_i f_i(x) d\mu = \xi_i, \quad i = 1, \dots, n$$

M DeGroot (1962).

Dessa forma, $E[\xi(X) | \xi] = \xi$. (vii)

Ou seja, a distribuição da Esperança da posteriori $\xi(X)$ dado ξ (priori), quando X ainda não é conhecido, é a própria distribuição a priori ξ .

Assim,

$$E\{U(\xi(X))|\xi\} \leq U\{\xi\} \text{ ou, simplificadamente,}$$

$$E(U([\theta | X])|\theta) \leq U(\theta)$$

$$E(U([\theta | X]) \leq U(\theta) \quad (\text{viii})$$

Aplicando a Variância, cônica, como a função de Incerteza U ,

$$E[V\{\theta | X\}] \leq V(\theta)$$

E, portanto

$$I\{X, P; V\} = V(\theta) - E(V[\theta | X]) \geq 0 \text{ quando } U \text{ é a variância.} \quad (\text{ix})$$

Ou seja, a Informação cresce na medida em que a esperança da nossa incerteza U , função cônica, tomada sobre a distribuição posterior $\theta | X$ diminui.

Gráficos ilustrativos das distribuições de θ a priori e a posteriori.

Nessa ilustração, a distribuição a priori $f(\theta)$ tem distribuição $\theta \sim N(0,1)$ e, a distribuição a posteriori $f(\theta | X_j = x_j)$, distribuição $\theta | X_j = x_j \sim N(1,0.25)$. Essa ilustração exemplifica um caso onde houve ganho de informação $I\{x, P; V\}$ com a realização do experimento $X_j = x_j$:

$$I\{x_j, P; V\} = V(\theta) - E(V[\theta | x_j]) = 1 - 0.25 = 0.75 \geq 0$$

Ou seja, mediante a realização do experimento $X_j = x_j$, a incerteza U sobre θ , medida pelo operador variância V , é menor na distribuição na distribuição a posteriori quando comparada com a distribuição a priori.

3) O Experimento de Máxima Variância de Bayes

Consideremos o problema onde existam 2 ou mais experimentos realizáveis representados por $X_1, X_2, \dots, X_n$ e um parâmetro θ cujo valor verdadeiro não é conhecido. Suponhamos que as funções de verossimilhança $L(\theta | x_j)$ sejam conhecidas para cada possível valor dos experimentos x_j , $j = 1, 2, \dots, n$, e que desejamos saber qual é o melhor experimento X_j a realizar dentre os n possíveis.

Já vimos em (i) e (ii), no capítulo 2, que as densidades posteriores de $\theta | x_j$ devem ser obtidas pelo operador de Bayes. A informação, tal como definida em (v, cap. 2) é dada por

$$I\{X_j; P, U\} = U(P) - E[U(P_{x_j})]$$

onde P representa a priori e P_{x_j} , a posteriori para o experimento j .

Com a propriedade (vi), supondo U côncava, temos que $I\{X_j; P, U\} = U(P) - E[U(P_{x_j})] \geq 0$.

Usando a variância como um operador de Incerteza (de natureza quadrática),

$$I\{X_j, P, U\} = V(\theta) - E[V(\theta | X_j)] = V(E[\theta | X_j])$$

sendo $I\{X_j, P, U\}$ uma medida da Variação provocada em $E[\theta | X_j]$.

Gráficos dos Histogramas das Densidades Posteriores das Habilidades θ tomados com base num item com parâmetros $a_j = 0.85$, $b_j = -0.26$ e $c_j = 0.21$, no Modelo da Teoria da Resposta ao Item de 3 parâmetros, detalhada no capítulo 7, considerando a priori $\theta \sim N(0, 1)$ e 1000000 amostras de θ .

Nesse caso, onde a priori $\theta \sim N(0,1)$ e os parâmetros do item segundo a Teoria da Resposta ao Item são $a_j = 0.85$, $b_j = -0.26$ e $c_j = 0.21$, temos que

$$E[\theta | X_j = 0] = E \left[\frac{L(\theta | X_j = 0) f(\theta)}{\int L(\theta | X_j = 0) f(\theta) d\theta} \right] = -0.4007982$$

$$E[\theta | X_j = 1] = E \left[\frac{L(\theta | X_j = 1) f(\theta)}{\int L(\theta | X_j = 1) f(\theta) d\theta} \right] = 0.2236248$$

$$P(X_j = 0 | \xi) = \int L(\theta | X_j = 0) f(\theta) d\theta = 0.357258$$

$$P(X_j = 1 | \xi) = \int L(\theta | X_j = 1) f(\theta) d\theta = 0.642742$$

onde ξ representa a priori $f(\theta)$ para a habilidade do examinando

A Variância $V(E[\theta | X_j])$, nesse caso, é a variância de dois eventos: (1) evento $E[\theta | X_j = 0]$ que ocorre com probabilidade $P(X_j = 0)$ e (2) evento $E[\theta | X_j = 1]$ que ocorre com probabilidade $P(X_j = 1)$.

Donde a Variância $V(E[\theta | X_j])$ é dada por

$$V(E[\theta | X_j]) = E\{E[\theta | X_j]^2\} - E^2\{E[\theta | X_j]\}$$

Com

$$\begin{aligned} E\{E[\theta | X_j]^2\} &= P(X_j = 0) \cdot (E[\theta | X_j = 0])^2 + P(X_j = 1) \cdot (E[\theta | X_j = 1])^2 = \\ &= 0.357258 \cdot (-0.4007982)^2 + 0.642742 \cdot (0.2236248)^2 = 0.08953194 \end{aligned}$$

$$\begin{aligned} E^2\{E[\theta | X_j]\} &= (P(X_j = 0) \cdot E[\theta | X_j = 0] + P(X_j = 1) \cdot E[\theta | X_j = 1])^2 = \\ &= [0.357258 \cdot (-0.4007982) + 0.642742 \cdot 0.2236248]^2 = 0.00000031 \end{aligned}$$

e assim,

$$V(E[\theta | X_j]) = 0.08953194 - 0.00000031 = 0.08953163$$

Adicionalmente temos, no mesmo exemplo,

$$V(\theta | X_j = 0) = \int (\theta - E[\theta | X_j = 0])^2 \cdot f(\theta | X_j = 0) d\theta = 0.8627928$$

$$V(\theta | X_j = 1) = \int (\theta - E[\theta | X_j = 1])^2 \cdot f(\theta | X_j = 1) d\theta = 0.9347546$$

e, portanto,

$$E[V(\theta | X_j)] = P(X_j = 0) \cdot V(\theta | X_j = 0) + P(X_j = 1) \cdot V(\theta | X_j = 1) =$$

$$= 0.357258 \cdot 0.8627928 + 0.642742 \cdot 0.934756 = 0.9090465$$

Observemos que $E[V(\theta | X_j)]$ será apresentado como Risco de Bayes associado à escolha do item j .

Futuramente iremos abordar o caso da escolha de um subconjunto J de experimentos, de tamanho k , entre uma gama maior de experimentos disponíveis (n). Desse modo, focaremos, inicialmente, a definição de Informação com base num único experimento tal qual abordado no texto desenvolvido até o momento.

Assim, nesta simplificação inicial, a Informação associada à realização do experimento X_j é uma medida de quanto esperamos reduzir nossa incerteza acerca de θ com a realização do experimento X_j . O experimento ideal reduziria nossa esperança da incerteza posterior a zero, caso esse, em que a Informação seria igual a $V(\theta)$. Um experimento ideal, nessa forma, revelaria o verdadeiro valor de θ . Para experimentos corriqueiros, que apenas reduzem nossa incerteza sobre θ , temos que o melhor experimento j^* dentre n experimentos realizáveis é, portanto, aquele que maximiza a informação $I\{X_j, P, U\} = V(\theta) - E[V(\theta | X_j)] = V(E[\theta | X_j])$, ou seja, o experimento j^* é aquele em que $E[\theta | X_j]$ é mais sensível aos valores amostrais.

Deste modo, o melhor experimento j é aquele tal que

$$j^* = \arg \max_j \{V(E[\theta | X_j])\}$$

Gráfico ilustrativo para o cálculo da Informação $I\{X_j, P, V\} = V\{E[\theta | X_j]\}$

O Gráfico acima ilustra o cálculo de $I\{X_j, P, V\} = V\{E[\theta | X_j]\}$ quando o experimento X_j ainda não foi realizado; o cálculo da esperança é feito para θ e o cálculo da variância, para X_j

4) O Experimento Ótimo à Luz da Teoria da Decisão

O intuito deste capítulo é estabelecer um paralelo entre o Experimento de Maior Informação de Bayes, visto no capítulo anterior, com o Experimento Ótimo escolhido à Luz da Teoria da Decisão. Veremos que no caso particular estudado, as duas abordagens são equivalentes quando a Função de Incerteza U (Experimento de Maior Informação de Bayes) e a Função de Perda L (Teoria da Decisão) forem convenientemente escolhidas.

Para formalizar o problema de decisão consideremos o espaço $(D, \Theta, \tilde{L}, \xi, S, F)$ onde D representa o espaço de decisões (no nosso caso, o espaço de experimentos "escolhíveis"); Θ representa o espaço paramétrico de θ , a habilidade de um examinando; $\tilde{L}$ representa a função de perda; ξ representa a distribuição a priori $f(\theta)$ e $\xi(X)$, a distribuição a posteriori $f(\theta | X)$; S representa o espaço amostral de x ; e F representa o conjunto de distribuições de $X | \theta_i$.

O espaço de regras de decisão é Δ , com δ função de S em D , levando $X \in S$ a $\delta(X) \in D$:

$$\delta: \begin{array}{l} S \rightarrow D \\ X \rightarrow \delta(X) \end{array}$$

A função de perda $\tilde{L}$ é uma função que associa um valor (perda) a cada possível valor de δ , θ e X :

$$\tilde{L}(\delta; \theta, X)$$

A escolha de um experimento X_d dentre n experimentos possíveis pode ser visto como uma regra de decisão δ que associa um subconjunto do espaço amostral $X_d \in S$ a $\delta(X_d) \in D$. Onde

X_d é um dos possíveis experimentos d , $d = (1, 2, \dots, n)$;

D é uma das possíveis decisões: $D = \{1, 2, \dots, n\}$.

A Regra de Decisão Ótima relativa a esse problema é

$$\delta^* = \inf_{\delta \in \Delta} E\{\tilde{L}(\delta; \theta, X)\} = \inf_{d \in D} E\{\tilde{L}(d; \theta, X_d)\}$$

Esse problema de decisão pode ser reescrito no espaço $(\Delta, \Theta \times S, \tilde{L}, \xi \times F)$ com

$$\tilde{L}(\delta; \theta, X) = L(\delta(X), \theta)$$

Donde sabemos que:

$$\delta^* = \inf_{\delta \in \Delta} E\{\tilde{L}(\delta; \theta, X)\} = \inf_{\delta \in \Delta} E\{L(\delta(X), \theta)\} \quad \text{M DeGroot (1970)}$$

Teorema: No contexto acima, isto é, $(\Delta, \Theta \times S, \tilde{L}, \xi \times F)$, onde $\Delta = \{\delta : \delta \text{ é função de } S \text{ em } D\}$, $\tilde{L} = L(\delta(X), \theta)$ é uma função de perda, ξ é uma medida de probabilidade em Θ , $F = \{f_X(\cdot | \theta) : \theta \in \Theta\}$ representa o conjunto de distribuições de $X | \theta_i$, temos:

$$\delta^*(X) = d_X^* \quad (\text{x})$$

onde d_X^* é uma decisão de Bayes contra a "priori" (posteriori) $\xi(X)$. d_X^* tem o significado da Decisão Ótima Estatística que, ante a regra de decisão definida, resulta na escolha do experimento X_d .

Tal teorema encontra-se provado (M DeGroot 1970) e a implicação, no nosso exemplo, numa análise pré-posteriori, é que devemos tomar a esperança de $\xi(X)$, $(E[\theta | X])$, como melhor estimativa de $\xi(X)$, quando X ainda não foi realizado.

Essa afirmação se sustenta quando consideramos que $\mu = E[\theta | X]$ minimiza $E[(\theta - c)^2]$, $c \in \mathbb{R}$, i.e.,

$$V(\theta | X) = E[(\theta - \mu)^2] = E[(\theta - E[\theta | X])^2] = \min_{c | X \in \mathbb{R}} E[(\theta - c)^2]$$

$\mu = E[\theta | X]$ é o Estimador de Bayes para perda quadrática (variância).

Ou seja, o melhor estimador pré-posteriori para $\xi(X)$ é $E[\xi(X)]$, considerando a perda quadrática.

Para o nosso problema, onde devemos escolher um experimento, a função de perda é dada, por $L(\delta(X_d), \theta) = (\theta - E[\theta | X_d])^2$ e, portanto,

$$L(d, \theta) = E[\tilde{L}(\delta, \theta, X)] = E[L(\delta(X_d), \theta)] = E[(\theta - E[\theta | X_d])^2] = V(\theta | X_d)$$

Dessa forma recaímos no problema de decisão de DeGroot no espaço $(d_X, \Theta \times S, L, \xi \times F)$ onde sabemos que a decisão ótima de Bayes é aquela que minimiza o Risco de Bayes

$$\rho_d = E[L(d, \theta)] = E[V(\theta | X_d)]$$

Ou seja, a Decisão Ótima d^* , que minimiza o Risco de Bayes, é aquela tal que

$$d^* = \arg \min_d \{E[V(\theta | X_d)]\}$$

Observemos que

$$\rho_d = E[V(\theta | X_d)] = P(X_d = 0) \cdot V(\theta | X_d = 0) + P(X_d = 1) \cdot V(\theta | X_d = 1)$$

Mas pela propriedade $V(\theta) = V(E[\theta | X_d]) + E[V(\theta | X_d)]$, a decisão d que minimiza o Risco de Bayes $\rho_d = E[L(d, \theta)] = E[V(\theta | X_d)]$ é aquela que maximiza a Função de Informação $I = V(E[\theta | X_d]) = V(\theta) - E[V(\theta | X_d)]$ pois $V(\theta)$ é conhecido.

O resultado é que a escolha do Experimento de Máxima Variância de Bayes, quando o operador sobre $\theta | X_d$ é a Esperança e a Função de Incerteza U é a variância, é equivalente à Decisão do Experimento Ótimo quando a Função de Perda é dada pela expressão $L(d, \theta, X_d) = (\theta - E[\theta | X_d])^2$ e o operador da Função de Perda é a Esperança.

Diante desse resultado, o problema sobre a escolha do melhor experimento pode ser realizado indiferentemente através do Experimento de Máxima Variância de Bayes ou através do Experimento Ótimo de Bayes com função de perda $L(d, \theta, X_d) = (\theta - E[\theta | X_d])^2$. Em termos abrangentes, a escolha do Experimento mais Informativo de Bayes equivalerá à Decisão de menor Risco de Bayes quando consistentemente definidos os operadores, a Função de Incerteza e a Função de Perda.

5) O problema da Escolha dos k Melhores Experimentos

Suponhamos desta forma que não desejamos apenas o melhor experimento, mas sim um conjunto J com k experimentos, os melhores possíveis a fim de reduzir nossa incerteza sobre o parâmetro θ de interesse. Suponhamos também que o mesmo experimento não possa ser realizado em duplicidade para a estimação do mesmo parâmetro θ . Esta suposição é relevante nesse momento pois consideramos que não devemos rerepresentar uma questão de uma prova a um mesmo examinando.

Relembrando o conceito de Informação, desejamos calcular uma medida de informação com base no subconjunto X_J ($X_J = X_{j1}, X_{j2}, \dots, X_{jk}$) dentre n ($X_1, X_2, \dots, X_n$) experimentos realizáveis.

A Informação contida nesses J experimentos é

$$I\{x_J, P, U\} = V(\theta) - E[V(\theta | x_J)] = V(E[\theta | x_J]) \geq 0$$

Assim, o melhor conjunto de experimentos J , de tamanho k , é aquele que maximiza a informação, é aquele tal que

$$J^* = \arg \max_J \{V(E[\theta | X_J])\} \quad J \in C_{n,k}$$

onde avaliam-se todos os subconjuntos J , combinação de k experimentos dentre n possíveis. Existem, desse modo, $C_{n,k}$ combinações de experimentos a serem avaliadas.

Segundo a Teoria da Resposta ao Item, considerando a independência condicional de $\theta | X_1, \theta | X_2, \dots, \theta | X_n$,

$$L(\theta | X_J) = \prod_{j=1}^k L(\theta | X_j)$$

E, portanto,

$$f(\theta | X_J) \propto L(\theta | X_J) \cdot f(\theta)$$

Novamente a informação está contida na verossimilhança $L(\theta | X_J) = \prod_{j=1}^k L(\theta | X_j)$

De modo análogo ao paralelo estabelecido com a Teoria da Decisão, a função de perda associada é

$$L(\theta, d_J, X_{J_d}) = (\theta - E[\theta | X_{J_d}])^2$$

onde d_J significa a decisão pela escolha do subconjunto X_{J_d} experimentos de tamanho k dentre $C_{n,k}$ combinações de experimentos realizáveis.

O Risco de Bayes da Decisão d_J é, portanto,

$$\rho_{d_J} = E[L(d_J, \theta)] = E[V(\theta | X_{d_J})]$$

Para o caso de dois experimentos dentre $C_{n,2}$:

$$\rho_{d_J} = \begin{cases} P(X_{d_{j1}}, X_{d_{j2}} = 0, 0) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}} = 0, 0) + \\ P(X_{d_{j1}}, X_{d_{j2}} = 0, 1) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}} = 0, 1) + \\ P(X_{d_{j1}}, X_{d_{j2}} = 1, 0) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}} = 1, 0) + \\ P(X_{d_{j1}}, X_{d_{j2}} = 1, 1) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}} = 1, 1) \end{cases}$$

onde $(X_{d_{j1}}, X_{d_{j2}})$ é uma combinação de 2 experimentos dentre n possíveis experimentos realizáveis.

Para um conjunto de tamanho k experimentos dentre $C_{n,k}$ possíveis combinações de experimentos realizáveis,

$$\rho_{d_J} = \begin{cases} P(X_{d_{j1}}, X_{d_{j2}}, \dots, X_{d_{jk}} = 0, 0, \dots, 0) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}}, \dots, X_{d_{jk}} = 0, 0, \dots, 0) + \\ P(X_{d_{j1}}, X_{d_{j2}}, \dots, X_{d_{jk}} = 0, 0, \dots, 1) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}}, \dots, X_{d_{jk}} = 0, 0, \dots, 1) + \\ \vdots + \\ P(X_{d_{j1}}, X_{d_{j2}}, \dots, X_{d_{jk}} = 1, 1, \dots, 1) \cdot V(\theta | X_{d_{j1}}, X_{d_{j2}}, \dots, X_{d_{jk}} = 1, 1, \dots, 1) \end{cases}$$

Lembrando que a Decisão Ótima de Bayes é aquela que minimiza o Risco de Bayes

$$d_J^* = \arg \min_{d_J} \{ \rho_{d_J} \} = \arg \min_{d_J} \{ V(E[\theta | X_J]) \}$$

Novamente, pela propriedade $V(\theta) = V(E[\theta | X_J]) + E[V(\theta | X_J)]$, verificamos que a Decisão Ótima de Bayes, considerando função de perda quadrática equivale à solução do Experimento de Máxima Variância de Bayes.

Para fins de exemplificação, consideremos um problema mais simples onde $\theta \in \{-1, 1\}$. O estado da natureza $\theta = -1$ representa a condição do paciente está saudável e $\theta = 1$ representa a condição do paciente está doente. Vamos assumir a priori $P(\theta = -1) = 0.4$ e $P(\theta = 1) = 0.6$. A fim de diminuir nossa incerteza sobre θ vamos analisar três ensaios clínicos diferentes representados pelas suas verossimilhanças, onde $X = 0$ representa o resultado negativo no ensaio clínico e $X = 1$, o resultado positivo:

Experimento (Ensaio Clínico) 1:

$$P_{X_1}(X_1 = 1 | \theta = -1) = 0.2 \quad P_{X_1}(X_1 = 0 | \theta = -1) = 0.8$$

$$P_{X_1}(X_1 = 1 | \theta = 1) = 0.9 \quad P_{X_1}(X_1 = 0 | \theta = 1) = 0.1$$

Experimento (Ensaio Clínico) 2:

$$P_{X_2}(X_2 = 1 | \theta = -1) = 0.3 \quad P_{X_2}(X_2 = 0 | \theta = -1) = 0.7$$

$$P_{X_2}(X_2 = 1 | \theta = 1) = 0.8 \quad P_{X_2}(X_2 = 0 | \theta = 1) = 0.2$$

Experimento (Ensaio Clínico) 3:

$$P_{X_3}(X_3 = 1 | \theta = -1) = 0.05 \quad P_{X_3}(X_3 = 0 | \theta = -1) = 0.95$$

$$P_{X_3}(X_3 = 1 | \theta = 1) = 0.7 \quad P_{X_3}(X_3 = 0 | \theta = 1) = 0.3$$

Inicialmente temos, a priori, que

$$E[\theta] = (-1) \cdot P(\theta = -1) + (1) \cdot P(\theta = 1) = 0.2 \text{ e}$$

$$V(\theta) = (-1)^2 \cdot P(\theta = -1) + (1)^2 \cdot P(\theta = 1) - (E[\theta])^2 = 0.96$$

Usando o operador de Bayes (iv) temos, no ensaio 1,

$$P(\theta = -1 | X_1 = 1) = \frac{P_{X_1}(X_1 = 1 | \theta = -1) \cdot P(\theta = -1)}{P_{X_1}(X_1 = 1 | \theta = -1) \cdot P(\theta = -1) + P_{X_1}(X_1 = 1 | \theta = 1) \cdot P(\theta = 1)} =$$

$$\frac{0.2 \cdot 0.4}{0.2 \cdot 0.4 + 0.9 \cdot 0.6} = 0.1290$$

$$P(\theta = 1 | X_1 = 1) = 0.8710$$

$$P(\theta = -1 | X_1 = 0) = \frac{P_{X_1}(X_1 = 0 | \theta = -1) \cdot P(\theta = -1)}{P_{X_1}(X_1 = 0 | \theta = -1) \cdot P(\theta = -1) + P_{X_1}(X_1 = 0 | \theta = 1) \cdot P(\theta = 1)} =$$

$$\frac{0.8 \cdot 0.4}{0.8 \cdot 0.4 + 0.1 \cdot 0.6} = 0.8421$$

$$P(\theta = 1 | X_1 = 0) = 0.1579$$

Logo,

$$E[\theta | X_1 = 1] = (-1) \cdot P(\theta = -1 | X_1 = 1) + (1) \cdot P(\theta = 1 | X_1 = 1) = 0.7419$$

$$E[\theta | X_1 = 0] = (-1) \cdot P(\theta = -1 | X_1 = 0) + (1) \cdot P(\theta = 1 | X_1 = 0) = -0.6842$$

$$P(X_1 = 1) = P_{X_1}(X_1 = 1 | \theta = -1) \cdot P(\theta = -1) + P_{X_1}(X_1 = 1 | \theta = 1) \cdot P(\theta = 1) = 0.62$$

$$P(X_1 = 0) = P_{X_1}(X_1 = 0 | \theta = -1) \cdot P(\theta = -1) + P_{X_1}(X_1 = 0 | \theta = 1) \cdot P(\theta = 1) = 0.38$$

$$V(\theta | X_1 = 1) = (-1)^2 \cdot P(\theta = -1 | X_1 = 1) + (1)^2 \cdot P(\theta = 1 | X_1 = 1) - (E[\theta | X_1 = 1])^2 =$$

$$= 0.4495$$

$$V(\theta | X_1 = 0) = (-1)^2 \cdot P(\theta = -1 | X_1 = 0) + (1)^2 \cdot P(\theta = 1 | X_1 = 0) - (E[\theta | X_1 = 0])^2 =$$

$$= 0.5319$$

A Variância e o Risco de Bayes associado ao experimento X_1 são dados por:

$$\nu_{X_1} = V(E[\theta | X_1]) = \left\{ \begin{array}{l} (E[\theta | X_1 = 1])^2 \cdot P(X_1 = 1) + \\ (E[\theta | X_1 = 0])^2 \cdot P(X_1 = 0) - \\ (E\{E[\theta | X]\})^2 \end{array} \right\} =$$

$$(0.7419)^2 \cdot 0.62 + (-0.6842)^2 \cdot 0.38 - (0.2)^2 = 0.4792$$

$$\rho_{X_1} = E[V(\theta | X_1)] = V(\theta | X_1 = 1) \cdot P(X_1 = 1) + V(\theta | X_1 = 0) \cdot P(X_1 = 0) =$$

$$= 0.4495 \cdot 0.62 + 0.5319 \cdot 0.38 = 0.4808$$

Para os demais casos, ensaios clínicos 2 e 3,

$$\nu_{X_2} = V(E[\theta | X_2]) = 0.24$$

$$\rho_{X_2} = E[V(\theta | X_2)] = 0.72$$

$$\nu_{X_3} = V(E[\theta | X_3]) = 0.3951$$

$$\rho_{X_3} = E[V(\theta | X_3)] = 0.5649$$

Note-se que nos três casos $\nu_{X_j} + \rho_{X_j} = V(E[\theta | X_j]) + E[V(\theta | X_j)] = 0.96 = V(\theta)$; Adicionalmente, o melhor experimento X_j , de Máxima Variância de Bayes, ou equivalentemente, o experimento de Mínimo Risco de Bayes sob perda quadrática é X_1 . ($\nu_{X_1} > \nu_{X_3} > \nu_{X_2}$ ou, analogamente, $\rho_{X_1} < \rho_{X_3} < \rho_{X_2}$).

Quando analisamos a escolha da combinação de 2 experimentos (ensaios clínicos, no exemplo) dentre 3 experimentos possíveis,

$$\nu_{X_1, X_2} = V(E[\theta | X_1, X_2]) = 0.5480$$

$$\rho_{X_1, X_2} = E[V(\theta | X_1, X_2)] = 0.4120$$

$$\nu_{X_1, X_3} = V(E[\theta | X_1, X_3]) = 0.6229$$

$$\rho_{X_1, X_3} = E[V(\theta | X_1, X_3)] = 0.3371$$

$$\nu_{X_2, X_3} = V(E[\theta | X_2, X_3]) = 0.5071$$

$$\rho_{X_2, X_3} = E[V(\theta | X_2, X_3)] = 0.4529$$

Note-se novamente que nos três casos $V(E[\theta | X_J]) + E[V(\theta | X_J)] = 0.96 = V(\theta)$; Adicionalmente, os melhores 2 experimentos $X_J = \{X_{j_1}, X_{j_2}\}$, de Máxima Variância de Bayes, ou equivalentemente, o conjunto de 2 experimentos de Mínimo Risco de Bayes sob perda quadrática é $\{X_1, X_3\}$, composto pelos 2 melhores experimentos individualmente. ($\nu_{X_1} > \nu_{X_3} > \nu_{X_2}$ ou, analogamente, $\rho_{X_1} < \rho_{X_3} < \rho_{X_2}$).

Consideremos o problema da escolha dos k melhores experimentos onde tomamos, ordenadamente, os 1º, 2º, 3º, ..., k º experimentos que revelem a maior informação sobre θ . Intuitivamente, diante de nenhuma nova informação sobre θ (não realização de nenhum experimento até o momento), obtemos que a escolha desses experimentos é a melhor escolha de tamanho k dentre n possíveis experimentos realizáveis, admitindo a não reposição de experimentos (não apresentação de um mesmo item a um examinando). Estamos admitindo, por hipótese, que a escolha dos J melhores experimentos é formada pela escolha ordenada dos k -ésimos $X = (X_{j_1}, \dots, X_{j_k})$ experimentos de Máxima Variância de Bayes (Mínimo Risco de Bayes), sucessivamente.

Tal afirmação decorre da propriedade de que quando consideramos perda quadrática para estimação de $\xi(X)|\xi$, o estimador é $E[\xi(X)|\xi]$. Desse modo, pela propriedade (vii) do capítulo 2,

$$E[\xi(X_1, X_2, \dots, X_{t-1})|\xi] = \xi$$

Usando a propriedade (x), do capítulo 4, para a Regra de Decisão, e considerando custo zero para a realização de cada experimento, temos que:

$$\delta^*(X_1, X_2, \dots, X_{t-1}) = d_{X_1, X_2, \dots, X_{t-1}}^*$$

Donde a "priori" a ser utilizada na escolha do t -ésimo experimento é $\xi(X_1, X_2, \dots, X_{t-1})$.

Assim, quando consideramos que $X_1, X_2, \dots, X_{t-1}$ ainda não foram realizados, devemos tomar a esperança de $\xi(X_1, X_2, \dots, X_{t-1})$ como "priori" para a escolha do t -ésimo experimento:

$$E[\xi(X_1, X_2, \dots, X_{t-1})] = E[\xi(X_1, X_2, \dots, X_{t-1})|\xi] = \xi$$

Ou seja, a "priori" a ser considerada para a escolha do t -ésimo experimento, dados que os experimentos anteriores ainda não foram realizados, é a própria priori inicial ξ de θ .

Assim, quando consideramos perda quadrática no espaço das regras de decisão $\delta(X)$, a posteriori $\xi(X)|\xi$, que deve ser usada para a escolha sucessiva dos 2º, 3º, ..., k º experimentos é a própria priori ξ , e, portanto, os k melhores experimentos podem ser obtidos via *ranking* e equivalem à solução da análise combinatória dos k melhores experimentos proposta no tópico anterior.

6) O Problema da Escolha Seqüencial dos k Melhores Experimentos

Para tratar do problema da seleção seqüencial de experimentos iremos adotar a visão da Teoria da Decisão pois vimos que para o caso de interesse os dois problemas (Decisão e Informação) são equivalentes.

No caso de decisões seqüenciais devemos tomar o espaço $(D, \Theta, L, \xi, S, F)$ onde D representa o espaço de decisões (no nosso o espaço de experimentos "escolhíveis"); Θ representa o espaço paramétrico de θ , a habilidade de um examinando; L representa a função de perda; ξ representa a distribuição a priori $f(\theta)$ e $\xi(X)$, a distribuição a posteriori $f(\theta|X)$; S representa o espaço amostral de x ; e F representa o conjunto de distribuições de $X|\theta_i$.

O problema de Decisão Estatística Seqüencial normalmente associa um custo de realização para cada Experimento X_j e utiliza uma regra de parada, função do custo (perda) acumulado até o instante t , onde a decisão de parar ou prosseguir com os experimentos é uma das possíveis decisões. No nosso caso iremos considerar que os custos para a realização de cada experimento são iguais a zero e a regra de parada é a realização de um número determinado (conhecido) k de Experimentos. Com isso, o problema de decisão seqüencial pode ser simplificado e atende melhor ao problema da realização de uma "Prova adaptada por Computador" onde normalmente desejamos prefixar o número de questões da Prova.

O espaço de regras de decisão é Δ , com δ função de S em D , levando $X \in S$ a $\delta(X) \in D$:

$$\delta: \begin{array}{l} S \rightarrow D \\ X \rightarrow \delta(X) \end{array}$$

A escolha de um experimento X_{d_t} dentre X_n experimentos possíveis pode ser visto como uma regra de decisão δ que associa, num instante t , um subconjunto do espaço amostral $X_{d_t} \in S$ a $\delta(X_{d_t}) \in D$.

Esse problema de decisão pode ser reescrito no espaço $(\Delta, \Theta \times S, \tilde{L}, \xi \times F)$ com

$$\tilde{L}(\delta; \theta, X) = L(\delta(X), \theta)$$

Donde sabemos que:

$$\delta^* = \inf_{\delta \in \Delta} E\{\tilde{L}(\delta; \theta, X)\} = \inf_{\delta \in \Delta} E\{L(\delta(X), \theta)\} \quad \text{M DeGroot (1970)}$$

Já vimos em (x), no capítulo 4, que a Decisão Ótima (aquela que minimiza o Risco de Bayes) é

$$\delta^*(X) = d_X^*$$

onde d_X^* é uma decisão de Bayes contra a "priori" $\xi(X)$.

Numa análise parcial-posteriori, o resultado é que devemos tomar $\xi(x_{1,\dots,t-1})$, ou seja, $f(\theta | x_{1,\dots,t-1})$ como "priori" para o problema da decisão ótima $d_{X_t}^*$.

Dessa forma, o problema de decisão seqüencial no instante t resume-se à escolha do melhor experimento $d_{X_t}^*$ contra a "priori" $\xi(x_{1,\dots,t-1})$.

Essa propriedade é conhecida como equivalência entre as formas normal e extensiva da Teoria da Decisão, onde a forma normal é o problema a "priori" com a regra de decisão $\delta^*(X)$ e a forma extensiva é o problema a "posteriori" com a decisão d_X^* .

Deste modo, a próxima decisão ótima tomará, sempre, tudo o que se conhece até o momento sobre o parâmetro θ , expressado por $\theta | x_1, \dots, x_{t-1}$, para então escolher o próximo experimento X_t , de menor Risco de Bayes, no nosso caso, admitindo experimentos sem reposição.

Observe-se que a escolha seqüencial de k questões de uma prova "Adaptativa" e a escolha "antecipada" de k questões de uma prova "Tradicional-Ótima" fundamentam-se na mesma base teórica, com a diferença que na prova "Adaptativa" usamos $\xi(x_{1,\dots,t-1})$ como "priori" no instante t ao passo em que usamos ξ , como "priori" para a escolha da t -ésima questão numa prova "Tradicional-Ótima"; a prova "Adaptativa", por conseguinte, parte de uma "priori" $\xi(x_{1,\dots,t-1})$ com mais informação, proporcionando uma escolha ótima com base nas informações acumuladas.

7) A Teoria da Resposta ao Item

O surgimento da Teoria da Resposta ao Item esteve intimamente ligada à necessidade de se criar um modelo teórico-estatístico que fosse capaz de comparar as habilidades (escores verdadeiros) dos examinandos submetidos à provas diferentes.

Os fundamentos da Teoria da Resposta ao Item consideram que as habilidades dos examinandos são características latentes e que não podem ser aferidas diretamente. Deste modo são utilizados métodos indiretos: os examinandos são submetidos a uma prova (conjunto de itens) e suas respostas são analisadas a fim de se estimar as habilidades de cada examinando individualmente.

Diferentemente da Teoria Clássica dos Testes (não confundir com Estatística Clássica), a **Teoria da Resposta ao Item** permite obter estimativas coerentes (comparáveis) para as habilidades de examinandos submetidos a provas diferentes entre si. Este é um grande diferencial entre a Teoria Clássica dos Testes e a Teoria da Resposta ao Item e que será fundamental para o desenvolvimento de provas otimizadas para uma população específica ou mesmo para cada examinando individualmente.

Para reforçar as vantagens da Teoria da Resposta ao Item serão apresentadas algumas ponderações críticas a respeito de comparações das habilidades no âmbito da Teoria Clássica dos Testes. Na ocasião estudava-se a comparação dos escores (uma estimativa da habilidade) de examinandos submetidos a "provas paralelas"¹. A conclusão desses estudos indicava a não aplicabilidade teórica da Teoria Clássica dos Testes na comparação das habilidades de examinandos submetidos a provas diferentes mediante transformações lineares dos escores pois provas diferentes possuem, na prática, dificuldades e fidedignidade diferentes e as relações entre os escores (exceto para provas verdadeiramente paralelas (e que não existem na prática)).

Em 1950 H Gulliksen (1950) definiu, no âmbito da Teoria Clássica dos Testes:

Duas provas podem ser consideradas formas paralelas quando, após a conversão para a mesma escala, suas médias, desvios padrão e correlações com todo e qualquer outro critério sejam iguais.

Em 1971, ainda no âmbito da Teoria Clássica dos Testes, W Angoff (1971) delimita a equivalência de provas ao apresentar as seguintes restrições (exigências):

- As provas devem medir a mesma característica ou habilidade;
- A equivalência estabelecida deve ser independente dos dados em particular utilizados para estabelecer esse princípio e deve ser aplicável em todas as situações parecidas;
- Os escores de duas provas, uma vez estabelecida sua equivalência, devem ser substituíveis entre si, e;
- A equivalência deve ser simétrica, ou seja, não deve fazer distinção entre a prova particular escolhida como base de referência.

¹ Provas Paralelas foi um conceito introduzido na Teoria Clássica dos Testes. Duas ou mais provas podem ser consideradas formas paralelas quando medem a mesma habilidade, possuem a mesma dificuldade e fidedignidade.

Angoff anota que a equivalência (ou conversibilidade) das provas só pode ser considerada onde as provas são paralelas, citando a definição de paralelismo de H Gulliksen (1950). Por último, Angoff enfatiza que apenas os escores de provas com coeficientes de erro iguais (fidedignidade² iguais) devam ser consideradas substituíveis.

As questões levantadas por Angoff foram reconsideradas e estendidas por F Lord (1977; 1980). A noção de escores substituíveis ganha nova dimensão com a introdução do conceito de equidade:

Os escores transformados y^* e observados x podem ser considerados "equivalentes" quando houver indiferença se um examinando responder a prova X ou Y (F Lord, 1977, pg 128).

De acordo com esse princípio:

- Torna-se inviável a tentativa de se estabelecer a equivalência entre provas que medem diferentes características ou habilidades (consistente com a primeira restrição de Angoff);
- A equivalência de escores com margens de erro desiguais não pode ser estabelecida (consistente com os escores substituíveis de Angoff);
- Não se pode estabelecer a equivalência de provas que refletem diferentes níveis de dificuldade.

Nas palavras de F Lord (1977, pg 128):

Se as provas X e Y têm dificuldades diferentes, a relação entre seus escores verdadeiros é necessariamente não linear devido aos efeitos de piso e de teto. Se duas provas têm uma relação não linear é implausível que elas sejam igualmente fidedignas em todos os grupos de examinandos. Isso leva à conclusão incômoda de que, a rigor, não se pode tornar equivalentes os escores observados em provas de dificuldades diferentes.

Todo esse panorama levou Lord a defender as vantagens teóricas dos modelos de resposta ao item em procedimentos que estabelecem a equivalência dos resultados de provas.

A Teoria da Resposta ao Item trata o problema de forma essencialmente diferente: o enfoque das análises desvincula-se das provas (Teoria Clássica dos Testes) e concentra-se nos Itens; se na Teoria Clássica dos Testes as estatísticas dos itens dependem da população dos examinandos e da prova à qual os itens pertencem, na Teoria da Resposta ao Item cria-se o conceito de que os parâmetros dos itens, obtidos no processo estatístico de "calibração" dos parâmetros dos itens (Valle, R. (2000)), são características próprias dos mesmos, e assim, os itens, representados por esses parâmetros, independem da população e da prova à qual pertencem ou venha a pertencer numa nova aplicação. Essa mudança no enfoque das análises permite a superação das grandes dificuldades teóricas na comparação das habilidades dos examinandos exigindo, apenas, que provas diferentes meçam as mesmas habilidades.

² O tema fidedignidade normalmente se desdobra em questões que incluem o tamanho e a dificuldade da prova; examinandos menos hábeis tendem a preferir provas menos fidedignas, onde a verdadeira habilidade dos examinandos é mascarada pela baixa ou alta dificuldade da prova; examinandos mais hábeis tendem a preferir provas mais fidedignas, que revelam mais sobre suas reais habilidades. Nesse sentido torna-se claro a busca por uma prova ótima, nem exageradamente fácil nem difícil, a fim de que elas revelem, ao máximo, as reais habilidades dos examinandos.

Desse modo, na Teoria da Resposta ao Item, os itens são posicionados numa mesma escala de habilidades; intuitivamente, se um examinando acertar 10 itens de dificuldade maior, esse examinando terá uma habilidade estimada maior que outro examinando que acertar 10 itens de dificuldade menor.

Gráfico Ilustrativo de 10 itens no modelo de Rasch
da Teoria da Resposta ao Item.

Existe uma grande gama de Modelos da Teoria da Resposta ao Item. Os modelos mais complexos podem incluir:

- **Multidimensionalidade da Habilidade:** essa modelagem incorpora a característica de que os itens medem mais de uma habilidade latente $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_k)$. A própria natureza de se medir uma característica latente é complexa: na prática é difícil de se afirmar inequivocamente a habilidade que um item está medindo; não é difícil encontrar casos reais concretos onde a habilidade supostamente medida por um item possa ser objeto de discussão; a inclusão de mais dimensões da habilidade suportada pelo modelo torna essa interpretação ainda mais complexa. Neste trabalho consideraremos que os itens medem apenas uma dimensão da habilidade.
- **Crédito Parcial:** essa modelagem incorpora a característica de que as respostas dos itens possam ser medidas em estágios hierárquicos: p.ex. 0,2 pontos se o examinando esquematizou o problema corretamente; 0,4 pontos se o examinando desenvolveu corretamente o raciocínio e; 0,4 pontos se os cálculos estão corretos. Essa modelagem pode ser muito adequada para provas dissertativas, contudo, diante do escopo do presente trabalho, que pretende analisar testes de múltipla escolha, os itens serão considerados como dicotômicos (certo ou errado).

Supondo, dessa forma, que a habilidade é unidimensional e que os itens são dicotômicos, ainda há uma certa gama de modelos da Teoria da Resposta ao Item que serão abordados no transcorrer desse capítulo.

Os principais fundamentos teóricos da Teoria da Resposta ao Item são:

- A probabilidade de um examinando qualquer acertar um item é modelada pela Curva Característica do Item. Essa Curva Característica do item é uma função da habilidade.

Considerando a modelagem de habilidade unidimensional e itens dicotômicos a Curva Característica do Item é uma função do tipo:

$$P(X_{ij} = 1 | \theta_i) = F_j(\theta_i) \quad i \text{ representa o examinando};$$

$$P(X_{ij} = 0 | \theta_i) = 1 - F_j(\theta_i) \quad j \text{ representa o item};$$

onde $F_j(\theta_i)$ é uma função monótona crescente a fim de acomodar o conceito de que a probabilidade de acerto é progressiva nos diversos níveis de habilidade. Adicionalmente, a probabilidade de acerto de cada item, dada a habilidade θ , é a mesma para todos os examinandos de mesma habilidade;

- Os itens possuem a propriedade de independência local, ou seja: apesar de X_{ij} , $j = 1, 2, \dots, J$ serem dependentes para um examinando i qualquer, $X_{ij} | \theta_i$, $j = 1, 2, \dots, J$ são independentes para todos e qualquer examinando i . Em outras palavras, conhecida a verdadeira habilidade θ_i de um examinando, a probabilidade de acerto do item j pelo examinando i não se altera em função dos acertos e erros de outros itens que esse mesmo examinando já tenha respondido. Caso fosse possível obter a verdadeira habilidade de um examinando i a aplicação de uma prova seria irrelevante pois sua habilidade já seria conhecida. Nos casos reais, como já foi dito, a habilidade é uma característica latente dos examinandos e a propriedade de independência local passa a ser de fundamental importância no cálculo da verossimilhança e, por conseguinte, para a estimação das habilidades dos indivíduos:

$$L(\theta_i | X_i = x_i, a, b, c) = \prod_{j=1}^J P(X_{ij} = x_{ij} | \theta_i, a_j, b_j, c_j)$$

onde a_j , b_j e c_j são, nesse exemplo, os parâmetros logísticos do item j , no modelo de 3 parâmetros da Teoria da Resposta ao Item.

- Outra característica dos modelos da Teoria da Resposta ao Item é que eles são não identificáveis. Essa característica está intimamente relacionada ao problema de construção de uma escala de habilidades: não há uma métrica de referência como uma régua milimetrada ou um termômetro padronizado para a habilidade. A solução genérica do modelo admite infinitas soluções onde, essas infinitas soluções dos parâmetros a , b e das habilidades θ_i , são analiticamente calculáveis mediante uma transformação linear de uma solução de referência; ou seja, uma vez construída uma escala de referência, as demais soluções do sistema não identificável podem ser obtidos mediante transformação linear. Desse modo, o problema da não identificabilidade dos modelos da Teoria da Resposta ao Item é costumeiramente resolvido através da imposição de uma escala de habilidades que determina a média e o desvio padrão das habilidades dos examinandos, tomando-se uma população qualquer de referência: a partir dessa imposição, os itens, bem como as habilidades dos examinandos, passam a ser referenciados nessa escala e a mudança para qualquer outra escala de habilidade pode ser facilmente obtida mediante uma transformação linear, resguardado o cuidado de que tanto os itens como os examinandos devem estar referenciados na mesma escala.
- A distribuição da habilidade posterior à aplicação de uma prova para um examinando i qualquer dentro da abordagem Bayesiana é obtida pela fórmula:

$$f(\theta_i | X_i = x_i, a, b, c) = \frac{f(\theta_i) \cdot L(\theta_i | X_i = x_i, a, b, c)}{\int_{-\infty}^{+\infty} f(\theta_i) \cdot L(\theta_i | X_i = x_i, a, b, c) \cdot d\theta_i}$$

onde :

- i. $f(\theta_i)$ é a distribuição a priori para o examinando i ;
- ii. $L(\theta_i | X_i = x_i, a, b, c)$ é a verossimilhança para θ_i considerando as respostas do examinando i ;
- iii. $f(\theta_i | X_i = x_i, a, b, c)$ é a distribuição posterior do examinando i considerando suas respostas na prova, e:
- iv. a, b, c são os vetores dos parâmetros dos itens para o caso do modelo de 3 parâmetros da Teoria da Resposta ao Item; nos modelos de Birnbaum e Rasch basta considerar os parâmetros pertinentes ao modelo.

Antes de introduzirmos os modelos logísticos da Teoria da Resposta ao Item, consideremos a seguinte modelagem para a probabilidade de acerto do item j em função da habilidade dos examinandos:

Gráfico Ilustrativo de um Item Ideal

Nessa modelagem estão implícitas várias simplificações:

1. O item é dicotômico, ou seja, $X_{ij}=1$ se o examinando i acerta o item j e $X_{ij}=0$ caso contrário;
2. A habilidade é unidimensional, ou seja, a modelagem pressupõe que apenas uma característica ou habilidade está sendo medida para cada examinando;
3. Além de unidimensional, essa característica ou habilidade está sendo medida igualmente para todos os examinandos indistintamente, ou seja, não há nenhum outro fator individual para todo e qualquer examinando influenciando na probabilidade de acerto senão a habilidade θ que se deseja medir;
4. A Curva Característica do Item j depende apenas de um parâmetro b_j ;

5. Nessa modelagem
$$\begin{cases} P(X_{ij}=1 | b_j) = 1, & \text{se } \theta_i \geq b_j & i = 1, 2, \dots, I & (I \text{ examinados}) \\ P(X_{ij}=1 | b_j) = 0, & \text{se } \theta_i < b_j & j = 1, 2, \dots, J & (J \text{ questões}) \end{cases}$$

Essa modelagem, contudo, pode ser vista como uma modelagem de um item ideal: bastaria um único item para certificar, inequivocamente, os examinandos considerados aptos (conjunto formado pelos examinandos i tais que $\theta_i \geq b_1$). Nesse processo seletivo hipotético (p.ex. um exame de habilitação veicular), b_1 seria exatamente a habilidade mínima necessária para os examinandos obterem o certificado de aptidão.

Para completar esse modelo, consideremos que os parâmetros b_j podem ser ordenados e referenciados numa escala única: $b = (b_{(1)}, b_{(2)}, \dots, b_{(J)})$. Assim, nesse contexto, passa a existir sentido em associar b_j como sendo o parâmetro de dificuldade do item j . Com efeito, um examinando i só acerta um item j se $\theta_i \geq b_j$; à medida que b_j aumenta, só os examinandos mais hábeis serão capazes de acertá-lo.

Uma questão oportuna de levantar é a independência condicional de $P(X_{ij} = 1 | b_j, \theta_i)$ para qualquer examinando i , fixado:

Ora, no modelo em questão, dados θ_i e b_j , obtém-se de imediato $P(X_{ij} = 1 | b_j, \theta_i) = 1$, se $\theta_i \geq b_j$ e $P(X_{ij} = 1 | b_j, \theta_i) = 0$, caso contrário; adicionar informações sobre outros itens respondidos é indiferente caso θ_i seja conhecido. Formalmente,

$$\begin{cases} P(X_{ij} = 1 | \underline{y}_i, b_j, \theta_i) = P(X_{ij} = 1 | b_j, \theta_i) = 1 & \text{se } \theta_i \geq b_j \\ P(X_{ij} = 1 | \underline{y}_i, b_j, \theta_i) = P(X_{ij} = 1 | b_j, \theta_i) = 0 & \text{se } \theta_i < b_j \end{cases}$$

$\underline{y}_i$ = respostas observadas em itens anteriores para o examinando i

Considerando a modelagem proposta e as respostas de um examinando i numa prova de J itens e aplicando o resultado da independência condicional, temos que a verossimilhança desse examinando i será:

$$L(\theta_i | \underline{b}, \underline{X}_i = \underline{x}_i) = \prod_{j=1}^J P(X_{ij} = x_{ij} | b_j, \theta_i)$$

Dada as características do modelo ora proposto a verossimilhança de um examinando i terá um dos seguintes perfis:

Gráficos Ilustrativos dos possíveis perfis da função de verossimilhança para as respostas de um examinando a um conjunto de Itens Ideais.

O primeiro gráfico refere-se a um examinando i que tenha errado todos os itens (θ_a , nesse caso é o valor mínimo dos parâmetros b_j da prova); o gráfico do centro refere-se a um examinando que tenha errado pelo menos um item e acertado ao menos algum outro (θ_a , nesse caso é o valor do parâmetro b_a do item mais difícil que o examinando acertou e θ_b é o valor do parâmetro b_b do item mais fácil que ele errou); o gráfico da esquerda refere-se a um examinando que tenha acertado todos os itens da prova (θ_b , nesse caso é o valor máximo dos parâmetros b_j da prova).

Note-se referentemente ao caso do gráfico central que, diante da modelagem ora proposta, $\theta_{ia} = \max(b_j | X_{ij} = 1)$ é necessariamente menor que $\theta_{ib} = \min(b_j | X_{ij} = 0)$. Em outras palavras, o modelo ilustrado não prevê que um examinando i possa acertar um item difícil e errar outro mais fácil; a verossimilhança para um caso desses seria igual a zero em todo o espaço de θ_i ($\theta_i \in]-\infty, +\infty[$).

A modelagem ora proposta permite, contudo, uma inferência precisa a respeito das habilidades dos examinandos independentemente da priori utilizada para as habilidades dos mesmos:

Examinando (i)	Resposta ao Item (1) [$b_{(1)}$]	Resposta ao Item (2) [$b_{(2)}$]	Resposta ao Item (3) [$b_{(3)}$]	Resposta ao Item (4) [$b_{(4)}$]	Resposta ao Item (5) [$b_{(5)}$]	Inferência Imediata
1	0	0	0	0	0	$\theta_1 < b_{(1)}$
2	1	0	0	0	0	$b_{(1)} \leq \theta_2 < b_{(2)}$
3	1	1	1	0	0	$b_{(3)} \leq \theta_3 < b_{(4)}$
4	1	1	1	0	0	$b_{(3)} \leq \theta_4 < b_{(4)}$
5	1	1	1	1	1	$\theta_5 \geq b_{(5)}$

Tabela 1: dados hipotéticos de examinandos submetidos a uma prova composta por Itens Ideais

Na tabela acima os itens (j) estão ordenados pelos parâmetros de dificuldade $b_{(j)}$ ($b_{(1)} < b_{(2)} < b_{(3)} < b_{(4)} < b_{(5)}$) e a inferência está generalizada para uma priori qualquer, desde que consistente com os padrões de resposta dos examinandos.

Nesse exemplo hipotético é possível estabelecer algumas inferências sobre as habilidades dos examinandos como, por exemplo, afirmar que $\theta_4 > \theta_2$, $\theta_5 > \theta_1$, etc.

O problema de se administrar provas diferentes a diferentes examinandos ganha nova dimensão. Para exemplificar os benefícios da nova forma de tratar o problema será apresentado um caso hipotético semelhante à tabela 1:

Examinando (<i>i</i>)	Resposta ao Item (1) [$b_{(1)}$]	Resposta ao Item (2) [$b_{(2)}$]	Resposta ao Item (3) [$b_{(3)}$]	Resposta ao Item (4) [$b_{(4)}$]	Resposta ao Item (5) [$b_{(5)}$]	Resposta ao Item (6) [$b_{(6)}$]	Resposta ao Item (7) [$b_{(7)}$]	Resposta ao Item (8) [$b_{(8)}$]	Inferência Imediata
1	0	0	0	0					$\theta_1 < b_{(1)}$
2		0	0	0	0				$\theta_2 < b_{(2)}$
3	1	0		0			0		$b_{(1)} \leq \theta_3 < b_{(2)}$
4			1		1	0		0	$b_{(5)} \leq \theta_4 < b_{(6)}$
5				1		1	0	0	$b_{(6)} \leq \theta_5 < b_{(7)}$
6	1	1				1	0		$b_{(6)} \leq \theta_6 < b_{(7)}$
7	1		1		1		1		$\theta_7 \geq b_{(7)}$
8		1		1		1		1	$\theta_8 \geq b_{(8)}$

Tabela 2: dados hipotéticos de examinandos submetidos a diferentes provas compostas por Itens Ideais

A tabela acima ilustra 8 examinandos fictícios que foram submetidos a subconjuntos diferentes de 4 itens dentre um total de 8 itens. Note-se que, nessa modelagem, as habilidades dos examinandos podem ser comparadas diretamente mesmo quando realizam provas diferentes. Note-se também que se itens suficientes forem escolhidos com b_j 's adequados é possível obter intervalos sem intersecção para as habilidades dos examinandos, o que permitiria ordená-los, inequivocamente, por habilidade, para fins de seleção, por exemplo.

Talvez o leitor, atendo-se aos examinandos 1 e 2 tenda a ver ressuscitadas as preocupações levantadas por Gulliksen, Angoff e Lord no âmbito da Teoria Clássica dos Testes. Ressalve-se, contudo, que dentro da Inferência Bayesiana as inferências estabelecidas para 1 e 2 refletem e melhor inferência diante do subconjunto dos dados amostrais colhidos para cada examinando. À medida que mais dados forem colhidos, melhor a inferência, e se a prova fosse suficientemente longa, com itens adequados, essa questão desapareceria mesmo com provas diferentes para diferentes examinandos, em contraste com a origem das questões levantadas por Gulliksen, Angoff e Lord.

Isto posto, torna-se mais relevante ressaltar o fato de que o modelo ora proposto não comporta que um examinando i erre um item mais fácil para depois acertar um item mais difícil. Este empecilho, que advém da idealização do modelo, torna relevante introduzir os modelos logísticos modernos da Teoria da Resposta ao Item, tratados a seguir.

Os modelos logísticos da Teoria da Resposta ao Item modelam a probabilidade de acerto do item em função da habilidade de acordo com uma curva logística. Uma das características da modelagem pela Teoria da Resposta ao Item é que sempre será possível obter as distribuições posteriores das habilidades dos examinandos submetidos a provas diferentes. Como já foi levantado na modelagem anterior, um certo cuidado sempre deverá estar presente nas inferências pois provas diferentes levarão a distribuições posteriores diferentes para um mesmo examinando i , examinando esse para o qual supõe-se que exista uma verdadeira habilidade θ_i que está sendo inferida indiretamente através dos acertos e erros dos itens. Na abordagem Bayesiana, contudo, esse fato deve ser tratado de forma natural e consistente: a distribuição posterior sempre reflete a informação a priori ponderada pela verossimilhança obtida através de dados amostrais; se a modelagem da Teoria da Resposta ao Item é plausível para inferir as habilidades de examinandos que realizaram a mesma prova então ela também será plausível para inferir as habilidades de outros examinandos que responderam a provas diferentes com o mesmo rigor teórico preservadas as seguintes restrições:

- As provas devem medir a mesma característica ou habilidade;
- Os parâmetros dos itens, assim como as habilidades (incluindo as distribuições a priori e a posteriori) dos examinandos, devem estar referenciadas na mesma escala;

○ Modelo Logístico de Rasch da Teoria da Resposta ao Item

A modelagem de Rasch considera que a probabilidade de acerto de cada item possa ser modelada pela seguinte curva logística:

$$P(X_{ij} = 1 | a, b_j, \theta_i) = \frac{1}{1 + e^{-Da(\theta_i - b_j)}}$$

Gráfico Ilustrativo de um Item no modelo de Rasch

A modelagem de Rasch considera apenas 1 parâmetro para cada item (b_j); o parâmetro a é um parâmetro de escala, fixo (igual) para todos os itens. Costumeiramente o valor de a é fixado em $a = 1.0$ quando o problema da não identificabilidade do modelo é resolvido através da imposição do valor 1.0, no exemplo, para desvio padrão da população de referência.

Nessa modelagem estão implícitas várias simplificações:

1. O item é dicotômico, ou seja, $X_{ij} = 1$ se o examinando i acerta o item j e $X_{ij} = 0$ caso contrário;
2. A habilidade é unidimensional, ou seja, a modelagem pressupõe que apenas uma característica ou habilidade está sendo medida para cada examinando;
3. Além de unidimensional, essa característica ou habilidade está sendo medida igualmente para todos os examinandos indistintamente, ou seja, não há nenhum outro fator individual para todo e qualquer examinando influenciando na probabilidade de acerto senão a habilidade θ que se deseja medir;
4. A curva característica do item j depende apenas de um parâmetro b_j e todos os itens têm o mesmo parâmetro a .

A modelagem de Rasch pode ser vista como uma extensão natural logística para a modelagem do item ideal, proposta no tópico anterior, admitindo-se que os itens não são

ideais a ponto de discriminar, inequivocamente, as habilidades dos examinandos num nível b_j da habilidade, característica do item j .

A Constante D da fórmula remete à origem da Teoria da Resposta ao Item onde chegou-se a modelar a Curva Característica do Item pela função densidade acumulada da distribuição Normal. Por conveniência matemática, a curva logística foi preferida na modelagem: quando $D = 1.7$ as Curvas Características do Item na abordagem Normal e Logística diferem de um valor menor que 0.01 em todo o espaço de habilidades ($\theta \in]-\infty, +\infty[$). Nesse caso, quando $D = 1.7$, o modelo é chamado de Modelo Logístico de Ogiva Normal da Teoria da Resposta ao Item; e, quando $D = 1.0$ o modelo é chamado genericamente de Modelo Logístico da Teoria da Resposta ao Item.

○ Modelo Logístico de Birnbaum da Teoria da Resposta ao Item

A modelagem de Birnbaum considera que a probabilidade de acerto de cada item possa ser modelada pela seguinte curva logística:

$$P(X_{ij} = 1 | a_j, b_j, \theta_i) = \frac{1}{1 + e^{-Da_j(\theta_i - b_j)}}$$

Gráficos Ilustrativos de dois Itens no modelo de Birnbaum

A modelagem de Birnbaum considera 2 parâmetros para cada item (a_j e b_j); o parâmetro b_j , por extensão do modelo de Rasch continua com o sentido de dificuldade do item e o parâmetro a_j ganha o sentido de poder de discriminação do item. Com efeito, tomando-se dois pontos θ_a e θ_b na vizinhança do ponto $\theta = b_j$, com $\theta_a < b_j$ e $\theta_b > b_j$ e, definindo

Discriminação $(\theta_a, \theta_b, X_j) = P(X_j = 1 | \theta_b) - P(X_j = 1 | \theta_a)$, quanto menor for a_j , menor será a variação na probabilidade de acerto (discriminação) numa vizinhança em torno de b_j e vice-versa. Quando $a_j \rightarrow +\infty$ a Curva Característica do Item se aproxima da curva do item ideal.

Nessa modelagem estão implícitas várias simplificações:

1. O item é dicotômico, ou seja, $X_{ij} = 1$ se o examinando i acerta o item j e $X_{ij} = 0$ caso contrário;
2. A habilidade é unidimensional, ou seja, a modelagem pressupõe que apenas uma característica ou habilidade está sendo medida para cada examinando;
3. Além de unidimensional, essa característica ou habilidade está sendo medida igualmente para todos os examinandos indistintamente, ou seja, não há nenhum outro fator individual para todo e qualquer examinando influenciando na probabilidade de acerto senão a habilidade θ que se deseja medir;
4. A curva característica do item j depende apenas de dois parâmetros: a_j e b_j , estando a_j associado ao sentido de discriminação do item e b_j ao sentido de dificuldade do item.

A modelagem de Birnbaum pode ser vista como uma extensão logística para a modelagem logística de Rasch. Com a modelagem de Birbaum estamos admitindo que cada um dos itens têm uma discriminação característica própria além da sua dificuldade.

○ Modelo Logístico com acerto casual (Guessing) da Teoria da Resposta ao Item

A modelagem com acerto casual (modelo de 3 parâmetros) considera que a probabilidade de acerto de cada item possa ser modelada pela seguinte curva logística:

$$P(X_{ij} = 1 | a_j, b_j, c_j, \theta_i) = c_j + (1 - c_j) \cdot \frac{1}{1 + e^{-Da_j(\theta_i - b_j)}}$$

Gráfico Ilustrativo de um Item no modelo de 3 parâmetros da TRI

A modelagem de 3 parâmetros da Teoria da Resposta ao Item considera 3 parâmetros para cada item (a_j , b_j e c_j); os parâmetros a_j e b_j continuam com o sentido atribuído na modelagem de Birnbaum e c_j ganha o sentido de "acerto casual". Com efeito, quando a habilidade $\theta \rightarrow -\infty$, $P(X_{ij} | a_j, b_j, c_j, \theta_i) \rightarrow c_j$, ou seja, c_j é a probabilidade do examinando i acertar o item quando sua habilidade θ_i é muito baixa, onde pressupõe-se que a única forma desse examinando acertar o item é através de uma escolha aleatória ("chute") dentre as alternativas apresentadas num teste de múltipla escolha.

Nessa modelagem estão implícitas várias simplificações:

1. O item é dicotômico, ou seja, $X_{ij} = 1$ se o examinando i acerta o item j e $X_{ij} = 0$ caso contrário;
2. A habilidade é unidimensional, ou seja, a modelagem pressupõe que apenas uma característica ou habilidade está sendo medida para cada examinando;
3. Além de unidimensional, essa característica ou habilidade está sendo medida igualmente para todos os examinandos indistintamente, ou seja, não há nenhum outro fator individual para todo e qualquer examinando influenciando na probabilidade de acerto senão a habilidade θ que se deseja medir;

4. A curva característica do item j depende apenas de três parâmetros: a_j , b_j e c_j , estando a_j associado ao sentido de discriminação do item e b_j ao sentido de dificuldade do item e c_j à característica de acerto casual do item.

O parâmetro de Acerto Casual é cercado de algumas controvérsias: primeiro discute-se que o acerto casual pode ser uma característica do examinando e não do item; a segunda discussão parte da modelagem em si, onde argumenta-se que pela natureza do parâmetro c_j ser uma característica mensurável para examinandos de baixíssima habilidade, então haveria pouca informação sobre o verdadeiro valor do parâmetro c_j . A primeira colocação pode ser contra-argumentada no momento em que os examinandos são estimulados (explicitamente ou implicitamente) a escolher uma resposta ao acaso quando os mesmos não sabem a verdadeira resposta num teste de múltipla escolha. Um caso explícito de estímulo à escolha ao acaso pode estar caracterizado quando é estabelecida a orientação de não deixar nenhuma resposta em branco no teste; um caso implícito é o próprio examinando tentar aumentar a sua nota (estimativa da habilidade) através do "chute" quando não há nenhuma punição para uma resposta errada num teste de múltipla escolha. Nesses casos faz sentido em se considerar o acerto casual como uma característica do item. A segunda colocação remete ao processo de calibração dos parâmetros dos itens que não será abordada neste trabalho, devendo o leitor consultar bibliografia como D Andrade (1998), R Klein (2003), R Valle (2000, 2001), Mislevy, J., Bock, D (1990). De todo modo, num processo de calibração Bayesiano essa preocupação perde sentido pois caso haja pouca informação amostral sobre os parâmetros c_j , então a distribuição a priori predominará para esses parâmetros. Essas distribuições a priori para os parâmetros c_j podem ser consistentemente definidas de acordo com o número de alternativas antes do processo de calibração dos itens.

Já foi comentado que os métodos de calibração dos itens segundo a Teoria da Resposta ao Item não são objeto de detalhamento do presente trabalho, contudo os dados obtidos a partir da calibração dos itens serão indispensáveis para o desenvolvimento dos capítulos seguintes visando a obter um método teórico-prático ótimo para a seleção dos itens que comporão uma prova específica. Deste modo, foram considerados os trabalhos de D Andrade (1998), R Klein (2003) e R Valle (2000, 2001) com o propósito da calibração dos itens de dois processos distintos: o primeiro remetendo ao 3º Exame de Certificação Qualificada realizado pela Associação Nacional dos Bancos de Investimento (ANBID) e o segundo ao processo de avaliação em matemática da 8ª série do Ensino Fundamental dos alunos do Serviço Social da Indústria (SESI), no ano de 2002.

No primeiro conjunto de dados (ANBID 2003), realizado em 20/07/2003 para 442 examinandos, foram escolhidas as 5 primeiras questões da prova de certificação com a finalidade de obter os parâmetros dos itens assim como as habilidades dos examinandos, calculados pelo BILOG. A finalidade dessa calibração dos itens pela TRI resume-se a comparar as habilidades dos examinandos estimadas pelo BILOG com as habilidades estimadas pelo processo de simulação bayesiano futuramente apresentado.

No segundo caso (SESI 2002), os alunos e itens de matemática da 8ª série foram selecionados e calibrados pela TRI num processo integral de avaliação que envolveu todo o universo de alunos do Estado de São Paulo em Português e Matemática nas 3ª, 5ª e 8ª séries do Ensino Fundamental e em Ciências nas 5ª e 8ª séries do Ensino Fundamental. Na avaliação, em Matemática, na 8ª série do Ensino Fundamental, especificamente, foram elaboradas 3 provas de 40 itens sendo que alguns desses itens eram comuns a duas ou a três provas. Essas 3 provas diferentes abordavam essencialmente a mesma habilidade e foram sorteadas aos 13.717 alunos que compareceram à avaliação.

Os itens em comum às provas permitiram a calibração de todos os itens numa mesma escala em conformidade com a Teoria da Resposta ao Item. Ao todo 90 itens diferentes de Matemática foram aplicados na 8ª série do Ensino Fundamental no SESI 2002.

Gráfico dos itens comuns às provas T, U e V de Matemática da 8ª do Ensino Fundamental Avaliação dos Alunos do SESI, 2002.

No primeiro caso (ANBID 2003), utilizando o programa comercial BILOG e a modelagem de 3 parâmetros da Teoria da Resposta ao Item, foram obtidos os seguintes parâmetros dos itens e, habilidades de 30 examinandos presentes sorteados aleatoriamente dentre os 442 examinandos presentes:

Item	$\hat{a}$	desvio padrão de a	$\hat{b}$	desvio padrão de b	$\hat{c}$	desvio padrão de c
0001	0,848930	0,258422	-0,259943	0,338163	0,208276	0,089154
0002	0,688038	0,208891	-0,152937	0,417558	0,222877	0,092473
0003	1,237744	0,434831	0,084595	0,238948	0,183074	0,079847
0004	0,636245	0,188988	-1,119507	0,495359	0,220962	0,094141
0005	0,927221	0,288577	-1,796837	0,498518	0,207291	0,091335

Tabela 3: parâmetros a , b e c e respectivos desvios-padrão, no modelo de 3 parâmetros da Teoria da Resposta ao Item, calculados para os 5 primeiros itens do exame de certificação ANBID 2003.

Examinando	Respostas	$\hat{\theta}$	desvio padrão de θ
000007D	BCCAD	0,801297	0,843041
000011F	DCAAD	-0,355353	0,782519
000020G	ACCAD	0,275241	0,831521
000025F	BCCBA	-0,250170	0,855070
000047E	BDCAA	-0,240675	0,844903
000055D	BCCAD	0,801297	0,843041
000092J	BAAAD	-0,269851	0,789342
000104B	CCCB	-0,118971	0,832748
000107H	CCCAC	-0,338208	0,838737
000128E	BDDCD	-0,609444	0,783089
000130C	BCCAD	0,801297	0,843041
000135B	BCCAD	0,801297	0,843041
000170D	CCCAD	0,275241	0,831521
000179K	BCCAD	0,801297	0,843041
000203D	ACAAD	-0,355353	0,782519
000254J	BCAAD	0,082965	0,802978
000255A	BCCAD	0,801297	0,843041
000262I	DACAD	-0,112932	0,823700
000276I	DBDAA	-1,153782	0,754656
000295B	BCCAD	0,801297	0,843041
000315D	BCCDD	0,385403	0,842542
000323C	BACAD	0,381751	0,834636
000339G	BCAAA	-0,478045	0,799251
000376B	BCAAD	0,082965	0,802978
000385C	BAAAD	-0,269851	0,789342
000406G	BACAD	0,381751	0,834636
000416J	BACAD	0,381751	0,834636
000444D	BCCAD	0,801297	0,843041
000452C	BABAD	-0,269851	0,789342
000463H	BABAD	-0,269851	0,789342

Tabela 4: respostas, habilidades θ e respectivos desvios-padrão calculados para os 30 examinandos sorteados aleatoriamente dentre os 442 examinandos presentes no exame de certificação ANBID 2003.

Gráficos das Curvas de Resposta ao Item dos 5 primeiros itens calibrados com 442 examinandos presentes no exame de certificação ANBID 2003.

Utilizando novamente o programa comercial BILOG, agora para a estimação dos itens de Matemática na 8ª série do Ensino Fundamental do SESI 2002, foram obtidos os seguintes parâmetros dos itens:

Item	$\hat{a}$	desvio padrão de $\hat{a}$	$\hat{b}$	desvio padrão de $\hat{b}$	$\hat{c}$	desvio padrão de $\hat{c}$
0001	0,493367	0,048585	0,264156	0,164287	0,129617	0,050316
0002	0,494281	0,036198	-1,397275	0,220960	0,152413	0,074465
0003	0,656179	0,061871	0,292201	0,115899	0,181445	0,040500
0004	0,458941	0,068711	1,291023	0,155811	0,186553	0,043790
0005	0,392948	0,053748	0,578515	0,262529	0,185217	0,063890
0006	0,279520	0,026421	-1,164538	0,470888	0,199908	0,088559
0007	0,617861	0,041333	0,761681	0,063906	0,105489	0,022843
0008	0,735503	0,142757	2,358819	0,173943	0,238117	0,015995
0009	1,131747	0,063839	0,693016	0,032294	0,188866	0,012750
0010	1,031067	0,098426	1,794677	0,063354	0,063795	0,008164
0011	0,109092	0,024361	9,443419	1,951343	0,080025	0,029670
0012	1,737807	0,176907	1,463528	0,037267	0,105928	0,007028
0013	1,080700	0,070239	1,290500	0,033708	0,173604	0,009430
0014	0,728578	0,074315	1,094939	0,070707	0,163796	0,023527
0015	0,874349	0,111710	1,250634	0,077089	0,348445	0,019887
0016	1,191994	0,081896	0,545207	0,037477	0,103734	0,016190
0017	1,574178	0,155058	1,088921	0,036799	0,221444	0,011390
0018	0,712116	0,050158	-0,072507	0,089338	0,093768	0,036384
0019	1,341619	0,180350	1,560848	0,057635	0,256837	0,011158
0020	0,673200	0,085319	0,139471	0,180561	0,429941	0,046188
0021	0,715881	0,051175	-0,900203	0,143272	0,149559	0,062179
0022	0,781366	0,072164	1,705218	0,057644	0,199823	0,012492
0023	1,195212	0,071310	1,671080	0,031067	0,148106	0,005651
0024	1,286593	0,082744	0,961594	0,030789	0,056341	0,009110
0025	0,427965	0,046720	-0,124878	0,255154	0,166460	0,068537
0026	1,066476	0,139818	1,827624	0,075582	0,203087	0,011294
0027	0,771562	0,125845	2,133933	0,128788	0,202337	0,015456
0028	1,041062	0,102207	1,217459	0,049000	0,161567	0,014618
0029	0,660295	0,032512	-1,781839	0,145280	0,141006	0,068365
0030	0,919421	0,091689	1,046807	0,054651	0,173345	0,019066
0031	0,727212	0,066868	1,279533	0,060651	0,085123	0,018135
0032	0,931014	0,041313	0,741297	0,029040	0,125737	0,011375
0033	0,522896	0,020717	-2,020800	0,136765	0,104420	0,054672
0034	0,892463	0,029682	-0,982224	0,053380	0,064935	0,028996
0035	0,707262	0,032279	-0,660946	0,090440	0,126080	0,039061
0036	0,459502	0,017198	-2,185127	0,114671	0,068518	0,037585
0037	0,504869	0,026837	-0,409659	0,125887	0,104636	0,042006
0038	0,984062	0,031450	-1,319818	0,049668	0,059346	0,029536
0039	0,643799	0,023058	-2,232272	0,094666	0,078630	0,042747
0040	0,854589	0,037122	-0,185284	0,054489	0,140492	0,024237
0041	1,142205	0,126186	1,628624	0,056517	0,183745	0,011442
0042	0,557044	0,080819	1,631573	0,106830	0,173526	0,030084
0043	0,522398	0,046199	0,609031	0,119631	0,171718	0,036992
0044	0,709947	0,049230	1,642380	0,047719	0,058419	0,010882
0045	1,246359	0,136471	1,253080	0,048884	0,265066	0,013992
0046	1,004794	0,108783	1,790861	0,065527	0,109781	0,010600
0047	0,047427	0,010980	17,587870	4,120259	0,013351	0,007303
0048	1,140378	0,138402	1,696178	0,062009	0,175265	0,011397
0049	0,588143	0,055593	0,173243	0,142148	0,150428	0,048492
0050	0,929479	0,092303	1,454490	0,055951	0,138288	0,014491
0051	0,971438	0,105839	1,651490	0,061831	0,137687	0,012691
0052	1,588452	0,150057	1,846751	0,039804	0,174898	0,005764
0053	0,871619	0,079899	0,996733	0,059134	0,184200	0,020462
0054	1,387393	0,167905	1,863333	0,060489	0,122715	0,007814
0055	0,834279	0,054071	-0,870724	0,109213	0,124220	0,053151
0056	1,198665	0,153583	1,681138	0,062253	0,243088	0,011794
0057	1,368259	0,125692	1,209117	0,037496	0,137943	0,011231
0058	0,648976	0,124131	2,511292	0,191420	0,169540	0,017737
0059	0,986934	0,117279	1,818890	0,071721	0,144966	0,011797
0060	0,991964	0,119827	1,360292	0,063568	0,287482	0,016866

0061	1,796632	0,204681	1,912899	0,041824	0,219107	0,005669
0062	0,046751	0,010701	13,823580	3,228270	0,021471	0,011531
0063	1,100874	0,090087	0,676354	0,047979	0,180141	0,019458
0064	0,882782	0,071929	0,434281	0,069486	0,188767	0,027318
0065	0,810860	0,092453	1,176142	0,071079	0,236469	0,022585
0066	0,958506	0,240294	2,855891	0,274965	0,182047	0,009378
0067	0,710845	0,040795	-1,215786	0,115358	0,103602	0,052406
0068	0,789092	0,047827	-0,614838	0,094024	0,100306	0,043023
0069	0,871425	0,074180	0,795611	0,059233	0,156571	0,022397
0070	0,772046	0,098767	1,784901	0,083746	0,144030	0,016788
0071	0,150154	0,025165	-0,848973	1,023565	0,272241	0,095308
0072	1,090451	0,092210	0,727405	0,050533	0,235029	0,018957
0073	0,538325	0,054091	0,053542	0,182947	0,192716	0,055925
0074	1,165454	0,156120	1,899933	0,073292	0,235066	0,010699
0075	1,244086	0,103460	1,424430	0,040319	0,090181	0,008975
0076	0,918962	0,059185	-0,260358	0,079278	0,137519	0,036349
0077	0,734795	0,047301	0,445803	0,062220	0,080037	0,023329
0078	0,597693	0,069698	1,761395	0,086863	0,097116	0,021046
0079	1,438180	0,133380	1,351768	0,038019	0,155267	0,009842
0080	0,942468	0,056667	1,376307	0,040410	0,023080	0,007240
0081	1,575579	0,122837	1,061785	0,031351	0,157240	0,010224
0082	0,728038	0,064835	0,212316	0,105415	0,191512	0,038793
0083	1,689588	0,208873	1,405610	0,042042	0,288540	0,010385
0084	1,294290	0,100104	0,978026	0,035313	0,141588	0,012468
0085	0,810795	0,061413	0,564407	0,063985	0,130582	0,024719
0086	1,225780	0,156355	2,023583	0,074630	0,142983	0,008448
0087	0,979064	0,106994	1,287396	0,055857	0,221658	0,016718
0088	0,720697	0,042721	-2,499473	0,146709	0,118976	0,063299
0089	1,439171	0,104211	0,806498	0,033551	0,170633	0,012757
0090	0,527697	0,048228	-1,329215	0,301559	0,261209	0,096352

Tabela 5: parâmetros a , b e c e respectivos desvios-padrão calculados pelo BILOG para os 90 itens da avaliação em Matemática da 8ª série do ensino Fundamental do SESI 2002.

8) Aplicação Prática na Escolha dos Melhores Itens de uma Prova

Vimos nos capítulos anteriores a fundamentação teórica para a escolha do melhor experimento a ser realizado estatisticamente. Os resultados obtidos são genéricos de modo que torna-se relevante a apresentação de um caso prático em que utilizaremos os resultados obtidos anteriormente.

Consideremos então o problema prático onde desejamos melhorar nossa incerteza acerca da habilidade (destreza, conhecimento) θ de um examinando e que dispomos de um certo conjunto de testes de múltipla escolha X_j , $j = 1, 2, \dots, J$ (dicotômicos) calibrados segundo a Teoria da Resposta ao Item.

De acordo com a TRI, modelo logístico de 3 parâmetros, as funções de verossimilhança associadas aos experimentos X_j são dadas pelas expressões

$$L(\theta | X_j = 1, a_j, b_j, c_j) = c_j + \frac{1 - c_j}{1 + e^{-Da_j(\theta - b_j)}}$$

$$L(\theta | X_j = 0, a_j, b_j, c_j) = 1 - c_j - \frac{1 - c_j}{1 + e^{-Da_j(\theta - b_j)}}$$

Gráficos das Verossimilhanças dado Acerto/Erro do primeiro item da ANBID 2003

Quando consideramos as respostas de um examinando a um certo conjunto de itens J , a verossimilhança das respostas aos experimentos (itens) terá um dos seguintes perfis, ao considerarmos o modelo logístico da Teoria da Resposta ao Item:

Gráficos das Verossimilhanças dos resultados dos 5 primeiros Experimentos ANBID 2003
Caso I: Erro nos 5 itens; Caso II: Erro nos 2º e 3º itens; Caso III: Acerto nos 5 itens

Logo, considerando a priori de θ , e obedecendo a abordagem do Experimento de Máxima Variância de Bayes, o melhor item j a se apresentar é aquele que maximiza

$$I\{X_j, P; V\} = V(\theta) - V(E[\theta | X_j]) \geq 0 \quad (\text{um único experimento})$$

$$I\{X_j, P; V\} = V(\theta) - V(E[\theta | X_j]) \geq 0 \quad (\text{mais de um experimento})$$

onde $f(\theta | X_j = 0) \propto L(\theta | X_j = 0) \cdot f(\theta)$ e $f(\theta | X_j = 1) \propto L(\theta | X_j = 1) \cdot f(\theta)$,

com $f(\theta)$ sendo a distribuição a priori para a habilidade θ e $L(\theta | x_j) = L(\theta | x_j, a_j, b_j, c_j)$ a função de verossimilhança (acerto e erro) referente cada um dos itens j . Os parâmetros de item a_j, b_j, c_j modelam a verossimilhança do acerto ao item segundo a curva logística de 3 parâmetros da Teoria da Resposta ao Item e são dados como conhecidos.

O fato de assumirmos que os parâmetros dos itens são conhecidos reflete uma simplificação do modelo pois, a rigor, segundo a modelagem completa da Teoria da Resposta ao Item, os parâmetros dos itens são incertos e, além de serem representados por uma distribuição (densidade de probabilidade), deveriam ser recalibrados (atualizados) a cada novo ensaio. Por hora, recorrendo-se à subjetividade do problema, iremos assumir que, antes de um novo julgamento a respeito do comportamento dos itens, os mesmos serão considerados como estáticos e as verossimilhanças dos acertos e erros ao item serão representadas pelos parâmetros logísticos a_j, b_j, c_j , certos e fixos, cujos valores refletem a Esperança a respeito dos mesmos. Note-se que o BILOG, programa comercial, na 3ª e última etapa, utiliza-se dessa mesma premissa para estimar as habilidades dos examinandos submetidos aos testes de múltipla escolha como verificamos na tabela 6 (Comparação entre BILOG e Simulação. $\theta_{ik} \sim N(0, 1)$; parâmetros a , b e c exatos. Dados ANBID 2003.).

Ressalte-se que a referência ao método de estimação do BILOG, citado no parágrafo anterior, não pretende justificar a simplificação em si. A simplificação do modelo se justifica subjetivamente à medida que um novo julgamento (reflexão) completo a respeito dos itens ainda não está disponível ante a apresentação de um novo item para algum examinando qualquer (eventualmente o mesmo examinando diante um novo item). Todos os cálculos, estimativas e resultados poderiam ser obtidos pelo método completo citado desde que as informações atualizadas estivessem disponíveis em cada momento, fato que no caso em estudo não se concretiza pois a reflexão psicométrica (reflexão humana) respeito dos parâmetros dos itens demanda um certo tempo para ser atestada pelos especialistas da área; dessa forma, durante esse ínterim, os parâmetros dos itens são dados como fixos (invariantes).

Há de se ponderar também a dificuldade matemática de se obterem os resultados exatos da distribuição posterior das habilidades $\theta_i | \vec{X}_{ij} = \vec{x}_{ij}$ ante as observações $\vec{X}_{ij} = \vec{x}_{ij}$ já que a verossimilhança, para vários itens respondidos por um examinando é um produtório de funções logísticas:

$$L(\theta_i | \vec{X}_{ij} = \vec{x}_{ij}) = \prod_{j_i} L(\theta_i | X_{ij} = x_{ij}) = \prod_{j_i} f(\theta_i | a_j, b_j, c_j)^{x_{ij}} \cdot [1 - f(\theta_i | a_j, b_j, c_j)]^{1-x_{ij}}$$

com

$$f(\theta_i | a_j, b_j, c_j) = c_j + \frac{1 - c_j}{1 + e^{-Da_j(\theta_i - b_j)}}$$

Iremos apresentar um método genérico de simulação para a distribuição posterior e, conseqüentemente, para a escolha dos melhores itens de uma prova.

Até o momento estamos admitindo que o parâmetro θ , o estado da natureza da habilidade, é uma variável contínua em $\mathfrak{R}$.

No caso prático entendemos que as habilidades de cada um dos examinandos é representada por θ . Nossa incerteza sobre a habilidade desses examinandos é representada por uma distribuição (contínua) de probabilidade em θ .

Na simulação estamos admitindo que essa distribuição contínua possa ser representada por um conjunto (grande) de amostras que simulem a distribuição original. Estamos admitindo que nossa incerteza sobre θ possa ser representada por esse conjunto de amostras simuladas.

Sabemos pelo operador de Bayes, no caso discreto (nossa simulação), que:

$$P(\theta | x_j) = \frac{L(\theta | x_j) \cdot P(\theta)}{\sum L(\theta | x_j) \cdot P(\theta)}$$

ou seja

$$P(\theta | x_j) \propto L(\theta | x_j) \cdot P(\theta)$$

Para simular a distribuição posterior devemos, portanto, ponderar a distribuição a priori pelo valor da verossimilhança.

Consideremos, então, o seguinte algoritmo:

- 1) Para cada valor de θ_{ik} simulado da distribuição a priori
- 2) Avaliar $L(\theta_{ik} | x_j = 1)$
- 3) Avaliar $u_{ik} \sim U[0, 1]$
- 4a) Se $u_{ik} \leq L(\theta_{ik} | x_j = 1)$ então a amostra θ_{ik} é selecionada para representar a distribuição posterior $P(\theta | x_j = 1)$
- 4b) Se $u_{ik} \geq L(\theta_{ik} | x_j = 1)$ então a amostra θ_{ik} é selecionada para representar a distribuição posterior $P(\theta | x_j = 0)$

Deste modo constituímos um algoritmo estocástico, por simulação, onde

$$P_{simulada}(\theta | x_j) \xrightarrow{n_{simulação} \rightarrow \infty} \frac{L(\theta | x_j) \cdot P(\theta)}{\sum L(\theta | x_j) \cdot P(\theta)}$$

Ou seja, se o tamanho da simulação for adequado, podemos utilizar o algoritmo para representar as distribuições posteriores completas.

Adicionalmente, neste algoritmo de simulação, obtemos estimadores:

$$p = \frac{\text{count}(u_{ik} \leq L(\theta_{ik}, x_d = 1))}{n_{\text{simulação}}} = E[u_{ik} \leq L(\theta_{ik} | x_d = 1)] \rightarrow P(X_d = 1 | \xi) \text{ e}$$

$$q = \frac{\text{count}(u_{ik} > L(\theta_{ik}, x_d = 0))}{n_{\text{simulação}}} = E[u_{ik} > L(\theta_{ik} | x_d = 0)] \rightarrow P(X_d = 0 | \xi)$$

onde

p é um estimador da probabilidade de um examinando i acertar o item d , considerando a incerteza a priori sobre θ_i e q , a respectiva probabilidade do mesmo examinando errar o item d , naquele momento. $L(\theta_{ik} | x_d = 0)$ e $L(\theta_{ik} | x_d = 1)$ são as verossimilhanças do erro/acerto do item x_d .

Com essa simulação temos a representação completa das distribuições a priori, e das distribuições posteriores para cada um dos possíveis experimentos j . No problema de decisão seqüencial, a cada momento, teremos, do mesmo modo, uma simulação da habilidade atualizada para a escolha da próxima questão da prova.

```
> mean(temp)
[1] -0.5622561
> var(temp)
[1] 0.684341
> summary(temp)
   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
-3.949000 -1.103000 -0.545700 -0.562300 -0.006417  2.831000
> length(temp)
[1] 79661
```

```
> mean(temp)
[1] 0.3783899
> var(temp)
[1] 0.8596014
> summary(temp)
   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
-3.8920 -0.2187  0.3951  0.3784  0.9933  4.4070
> length(temp)
[1] 120339
```

Os gráficos acima ilustram as distribuições posteriores dado erro/acerto a um item fictício de parâmetros $a = 1.0$, $B = 0.0$, $c = 0.2$ considerando a priori $\theta \sim N(0, 1)$.

O Risco de Bayes na simulação é

$$\rho_{X_j} = q \cdot V(\theta_{ik} | X_j = 0) + p \cdot V(\theta_{ik} | X_j = 1)$$

Consideremos agora o caso onde um examinando responde a um certo conjunto de dois itens de uma prova:

$$P(\theta | x_j) \propto L(\theta | x_j) \cdot P(\theta)$$

com

$$L(\theta | x_j) = \prod_{i \in J} L(\theta | x_j) = \prod_{i \in J} f(\theta | a_j, b_j, c_j)^{x_j} \cdot [1 - f(\theta | a_j, b_j, c_j)]^{1-x_j}$$

Novamente teremos

$$P_{simulada}(\theta | x_j) \xrightarrow{n_{simulação} \rightarrow \infty} \frac{L(\theta | x_j) \cdot P(\theta)}{\sum L(\theta | x_j) \cdot P(\theta)}$$

$$p_{00} = \frac{\text{count}(u_{ik} \leq L(\theta_{ik} | x_{J_d} = 0,0))}{n_{simulação}} = E[u_{ik} \leq L(\theta_{ik} | x_{J_d} = 0,0)] \rightarrow P(X_{J_d} = 0,0 | \xi)$$

$$p_{01} = \frac{\text{count}(u_{ik} \leq L(\theta_{ik} | x_{J_d} = 0,1))}{n_{simulação}} = E[u_{ik} \leq L(\theta_{ik} | x_{J_d} = 0,1)] \rightarrow P(X_{J_d} = 0,1 | \xi)$$

$$p_{10} = \frac{\text{count}(u_{ik} \leq L(\theta_{ik} | x_{J_d} = 1,0))}{n_{simulação}} = E[u_{ik} \leq L(\theta_{ik} | x_{J_d} = 1,0)] \rightarrow P(X_{J_d} = 1,0 | \xi)$$

$$p_{11} = \frac{\text{count}(u_{ik} \leq L(\theta_{ik} | x_{J_d} = 1,1))}{n_{simulação}} = E[u_{ik} \leq L(\theta_{ik} | x_{J_d} = 1,1)] \rightarrow P(X_{J_d} = 1,1 | \xi)$$

E o Risco de Bayes,

$$\rho_{X_j} = p_{00} \cdot V(\theta_{ik} | X_j = 0,0) + p_{01} \cdot V(\theta_{ik} | X_j = 0,1) + p_{10} \cdot V(\theta_{ik} | X_j = 1,0) + p_{11} \cdot V(\theta_{ik} | X_j = 1,1)$$

Com a finalidade de validar o método de simulação proposto tomamos as primeiras 5 questões do exame aplicado para certificação da ANBID. Numa primeira etapa, os itens foram calibrados pelo BILOG, assim como foram obtidas as respectivas estimativas de habilidade dos examinandos considerando esses 5 itens. Dentre os 442 examinandos que realizaram o exame, foram escolhidos, aleatoriamente, 30 examinandos para a comparação. A simulação da distribuição posterior para esses 30 examinandos foi obtida pelo método sugerido, utilizando a distribuição posterior do examinando dado acerto/erro ao primeiro item como a "priori" para a obtenção da distribuição posterior do segundo item e assim por diante até o quinto item selecionado. Os resultados encontram-se na tabela 6 a seguir:

Examinando	Simulação		Estimação BILOG		Comparação	
	NDEF	RESP	N-POST	MEAN(θ_{ik}) SE(θ_{ik})	Teta SE(Teta)	MEAN(θ_{ik})-Teta SE(θ_{ik})/SE(Teta)
1 000007D	BCCAD		101114	0,803 0,848	0,801 0,843	0,002 1,006
2 000011F	DCAAD		26512	- 0,357 0,779	- 0,355 0,783	- 0,002 0,995
3 000020G	ACCAD		36116	0,264 0,831	0,275 0,832	- 0,011 0,999
4 000025F	BCCBA		3373	- 0,255 0,846	- 0,250 0,855	- 0,005 0,989
5 000047E	BDCAA		5408	- 0,223 0,837	- 0,241 0,845	0,018 0,990
6 000055D	BCCAD		101316	0,801 0,850	0,801 0,843	0,000 1,009
7 000092J	BAAAD		26589	- 0,274 0,788	- 0,270 0,789	- 0,004 0,998
8 000104B	CCCBD		11797	- 0,111 0,837	- 0,119 0,833	0,008 1,005
9 000107H	CCCAC		5333	- 0,338 0,850	- 0,338 0,839	0,000 1,014
10 000128E	BDDCD		11882	- 0,613 0,784	- 0,609 0,783	- 0,003 1,002
11 000130C	BCCAD		100997	0,805 0,848	0,801 0,843	0,004 1,005
12 000135B	BCCAD		101782	0,806 0,851	0,801 0,843	0,004 1,010
13 000170D	CCCAD		35979	0,279 0,830	0,275 0,832	0,003 0,999
14 000179K	BCCAD		101447	0,801 0,850	0,801 0,843	0,000 1,008
15 000203D	ACAAD		26792	- 0,362 0,779	- 0,355 0,783	- 0,006 0,996
16 000254J	BCAAD		45748	0,080 0,799	0,083 0,803	- 0,003 0,996
17 000255A	BCCAD		101618	0,804 0,848	0,801 0,843	0,003 1,006
18 000262I	DACAD		19142	- 0,115 0,819	- 0,113 0,824	- 0,002 0,994
19 000276I	DBDAA		6274	- 1,157 0,775	- 1,154 0,755	- 0,004 1,027
20 000295B	BCCAD		101587	0,804 0,850	0,801 0,843	0,003 1,009
21 000315D	BCCDD		24197	0,384 0,845	0,385 0,843	- 0,001 1,003
22 000323C	BACAD		39885	0,385 0,841	0,382 0,835	0,003 1,007
23 000339G	BCAAA		7785	- 0,469 0,807	- 0,478 0,799	0,009 1,010
24 000376B	BCAAD		45407	0,072 0,803	0,083 0,803	- 0,011 0,999
25 000385C	BAAAD		27019	- 0,270 0,788	- 0,270 0,789	- 0,001 0,999
26 000406G	BACAD		39992	0,379 0,831	0,382 0,835	- 0,003 0,996
27 000416J	BACAD		40225	0,386 0,840	0,382 0,835	0,005 1,007
28 000444D	BCCAD		101388	0,804 0,850	0,801 0,843	0,003 1,008
29 000452C	BABAD		26900	- 0,272 0,790	- 0,270 0,789	- 0,003 1,001
30 000463H	BABAD		26898	- 0,270 0,794	- 0,270 0,789	0,000 1,006
MÉDIA				0,119 0,823	0,119 0,820	0,000 1,003

Tabela 6: Comparação entre BILOG e Simulação. $\theta_{ik} \sim N(0,1)$, parâmetros a , b e c exatos. Dados ANBID 2003. (Gabarito: BCCAD)

Na tabela anterior, a coluna Examinando representa o Examinando sorteado aleatoriamente para o estudo (NDEF representa o número de inscrição desse examinando); N-POST indica o número final de amostras que representam a distribuição posterior, dado os acertos e erros às questões selecionadas, obtidas via simulação; $MEAN(\theta_{ik})$ representa a estimativa pontual da habilidade do examinando na simulação; $SE(\theta_{ik})$ representa a estimativa pontual da incerteza (desvio padrão) sobre o examinando na simulação dada suas as respostas às questões. Teta e SE(Teta) representam as estimativas da habilidade e respectiva incerteza (desvio padrão) obtidas pelo BILOG. As últimas duas colunas estabelecem uma comparação entre as estimativas obtidas pelo método de simulação sugerido e os valores obtidos pelo BILOG. Notemos que o método de simulação proposto confere estimativas plenamente compatíveis com os valores estimados pelo BILOG; a rigor, o método de simulação é tão mais perfeito quanto maior for o tamanho da simulação da priori utilizada para representar a habilidade dos examinandos.

Note-se, eticamente, que os 30 examinandos foram escolhidos aleatoriamente dentre os 442 examinandos apenas para validar o algoritmo de simulação; a prova real, em lápis e papel, constou de 80 itens dos quais foram selecionados apenas 5 para fins ilustrativos. A inferência completa acerca da certificação ou não dos examinandos no procedimento de certificação deve considerar tão e somente apenas a prova completa tal qual regida em edital.

Escolha dos Itens de uma Prova por Simulação

A partir do desenvolvimento teórico-prático ora apresentado, a escolha dos melhores itens, itens de Máxima Variância de Bayes, via simulação torna-se então imediato:

Considerando os resultados anteriores, temos como obter uma simulação de

$$\theta | X_d = 0$$

$$\theta | X_d = 1$$

e, portanto, suas variâncias na simulação. Do mesmo modo, na mesma simulação, podemos obter os estimadores

$$p = E[u_{ik} \leq L(\theta_{ik} | x_d = 1)] \rightarrow P(X_d = 1 | \xi) \text{ e}$$

$$q = E[u_{ik} > L(\theta_{ik} | x_d = 1)] \rightarrow P(X_d = 0 | \xi)$$

Para escolher os melhores itens de uma prova basta, deste modo, tomar a(s) questão(ões) de Mínimo Risco de Bayes com os dados simulados à luz da teoria da decisão, escolhendo-se o(s) item(ns) tal(tais) que o risco

$$\rho_d = E[V(\theta | X_d)] = P(X_d = 0) \cdot V(\theta | X_d = 0) + P(X_d = 1) \cdot V(\theta | X_d = 1)$$

$$\rho_{d_{\text{simulação}}} = q \cdot V(\theta_{\text{simulado}} | X_d = 0) + p \cdot V(\theta_{\text{simulado}} | X_d = 1), \text{ na simulação,}$$

seja(m), ordenadamente, o(s) melhor(es) itens, ordenados pelo Risco de Bayes, dentre o conjunto de questões disponíveis.

Consideremos o conjunto de itens do Banco de Itens "SESI", constituído a partir da calibração dos itens da 8ª série em Matemática para a rede SESI em 2002.

Suponhamos que a habilidade de um novo examinando que realizará uma prova possa ser representada por uma distribuição referenciada na mesma escala de habilidades desse banco de itens. No caso suponhamos $\theta \sim N(0, 1)$ na escala "SESI". A escolha da prova ideal para esse novo examinando, considerando a seleção ótima dos itens do banco SESI, deverá tomar, como visto anteriormente, os k Itens de Menor Risco de Bayes. Deste modo, considerando o método de simulação ora proposto e aplicado-o nos dados da tabela 7 a seguir, basta-nos selecionar, os k itens de Menor Risco de Bayes, via Simulação. Para uma prova de 30 itens e priori $\theta \sim N(0, 1)$ segue que a Prova Ideal é obtida mediante a seleção dos 30 primeiros itens da tabela a seguir:

Ordem	Rótulo do Item j	a	b	c	Esperança de Acerto	Risco de Bayes
01	0024	1,286593	0,961594	0,056341	31,95%	0,834207
02	0016	1,191994	0,545207	0,103734	44,05%	0,844569
03	0089	1,439171	0,806498	0,170633	42,01%	0,859678
04	0081	1,575579	1,061785	0,157240	35,55%	0,869519
05	0084	1,294290	0,978026	0,141588	37,87%	0,874468
06	0038	0,984062	-1,319818	0,059346	76,33%	0,878279
07	0034	0,892463	-0,982224	0,064935	70,05%	0,883594
08	0076	0,918962	-0,260358	0,137519	61,24%	0,884653
09	0080	0,942468	1,376307	0,023080	26,70%	0,886681
10	0012	1,737807	1,463528	0,105928	24,09%	0,886841
11	0057	1,368259	1,209117	0,137943	33,12%	0,886861
12	0009	1,131747	0,693016	0,188866	47,11%	0,888513
13	0063	1,100874	0,676354	0,180141	47,03%	0,889485
14	0075	1,244086	1,424430	0,090181	27,18%	0,892603
15	0017	1,574178	1,088921	0,221444	40,15%	0,897584
16	0040	0,854589	-0,185284	0,140492	59,93%	0,897889
17	0032	0,931014	0,741297	0,125737	43,80%	0,898765
18	0055	0,834279	-0,870724	0,124220	69,62%	0,901893
19	0068	0,789092	-0,614838	0,100308	64,46%	0,902477
20	0079	1,438180	1,351768	0,155267	31,76%	0,904106
21	0072	1,090451	0,727405	0,235029	49,78%	0,906980
22	0064	0,882782	0,434281	0,188767	52,87%	0,911883
23	0085	0,810795	0,564407	0,130582	47,87%	0,912739
24	0077	0,734795	0,445803	0,080037	47,27%	0,913728
25	0069	0,871425	0,795611	0,156571	45,55%	0,916006
26	0018	0,712116	-0,072507	0,093768	55,67%	0,916162
27	0028	1,041062	1,217459	0,161567	37,98%	0,916971
28	0010	1,031067	1,794677	0,063795	22,79%	0,918073
29	0013	1,080700	1,290500	0,173604	37,43%	0,920986
30	0030	0,919421	1,046807	0,173345	42,59%	0,922595
31	0035	0,707262	-0,660946	0,126080	65,31%	0,923645
32	0067	0,710845	-1,215786	0,103602	71,86%	0,925454
33	0021	0,715881	-0,900203	0,149559	69,42%	0,925972
34	0053	0,871619	0,996733	0,184200	44,60%	0,928561
35	0050	0,929479	1,454490	0,138288	34,48%	0,929535
36	0031	0,727212	1,279533	0,085123	36,14%	0,931433
37	0082	0,728038	0,212316	0,191512	56,76%	0,932405
38	0023	1,195212	1,671080	0,148106	29,13%	0,932926
39	0045	1,246359	1,253080	0,265066	43,36%	0,933420
40	0046	1,004794	1,790861	0,109781	26,99%	0,934238
41	0044	0,709947	1,642380	0,058419	30,16%	0,935888
42	0051	0,971438	1,651490	0,137687	31,44%	0,936271
43	0054	1,387393	1,863333	0,122715	22,97%	0,937593
44	0087	0,979064	1,287396	0,221658	42,27%	0,938167
45	0041	1,142205	1,628624	0,183745	33,17%	0,940237
46	0014	0,728578	1,094939	0,163796	43,84%	0,941373
47	0007	0,617861	0,761681	0,105489	45,77%	0,941543
48	0083	1,689588	1,405610	0,288540	40,44%	0,941946
49	0048	1,140378	1,696178	0,175265	31,74%	0,942311
50	0003	0,656179	0,292201	0,181445	55,46%	0,942383
51	0029	0,660295	-1,781839	0,141006	78,14%	0,943632
52	0059	0,986934	1,818890	0,144966	29,64%	0,944236
53	0065	0,810860	1,176142	0,236469	46,78%	0,947525
54	0039	0,643799	-2,232272	0,078630	80,70%	0,947695
55	0049	0,588143	0,173243	0,150428	55,51%	0,947919
56	0019	1,341619	1,560848	0,256837	38,18%	0,948081
57	0088	0,720697	-2,499473	0,118976	85,69%	0,948227
58	0070	0,772046	1,784901	0,144030	33,83%	0,951166
59	0052	1,588452	1,846751	0,174898	26,32%	0,951565
60	0060	0,991964	1,360292	0,287482	46,18%	0,951784
61	0086	1,225780	2,023583	0,142983	24,67%	0,952175
62	0056	1,198665	1,681138	0,243088	36,95%	0,953480
63	0026	1,066476	1,827624	0,203087	33,44%	0,954719
64	0037	0,504869	-0,409659	0,104636	59,62%	0,954759
65	0078	0,597693	1,761395	0,097116	34,39%	0,956576
66	0022	0,781366	1,705218	0,199823	38,75%	0,957537
67	0073	0,538325	0,053542	0,192716	59,02%	0,958776
68	0033	0,522896	-2,020800	0,104420	75,86%	0,959839
69	0001	0,493367	0,264156	0,129617	53,73%	0,960068
70	0015	0,874349	1,250634	0,348445	53,06%	0,961049
71	0043	0,522398	0,609031	0,171718	52,45%	0,961352
72	0002	0,494281	-1,397275	0,152413	71,04%	0,962660
73	0074	1,165454	1,899933	0,235066	34,41%	0,963433
74	0090	0,527697	-1,329215	0,261209	74,79%	0,965439
75	0036	0,459502	-2,185127	0,068518	74,17%	0,965946
76	0020	0,673200	0,139471	0,429941	70,29%	0,966077
77	0061	1,796632	1,912899	0,219107	28,72%	0,966136
78	0042	0,557044	1,631573	0,173526	42,07%	0,967345
79	0027	0,771562	2,133933	0,202337	35,15%	0,969488
80	0025	0,427965	-0,124878	0,166460	59,37%	0,971173
81	0004	0,458941	1,291023	0,186553	48,13%	0,974999
82	0058	0,648978	2,511292	0,169540	32,05%	0,975825
83	0008	0,735503	2,358819	0,238117	37,00%	0,978176
84	0005	0,392948	0,578515	0,185217	54,82%	0,978781
85	0066	0,958506	2,855891	0,182047	25,12%	0,985788
86	0006	0,279520	-1,164538	0,199908	66,34%	0,988776
87	0071	0,150154	-0,848973	0,272241	65,94%	0,998449
88	0011	0,109092	9,443419	0,080025	32,27%	1,000003
89	0047	0,047427	17,587870	0,013351	31,20%	1,001269
90	0062	0,046751	13,823580	0,021471	35,79%	1,001351

Tabela 7: *Ranking* dos itens SESI 2002 ordenados de acordo com o Risco de Bayes considerando a priori $\theta \sim N(0, 1)$, obtida via simulação.

Prova Ideal para a cidade de Sertãozinho

Este tópico visa a explorar o conceito de informação e o conceito subjetivo da informação de Bayes. Para tanto foram extraídas algumas informações genéricas e específicas referentes a um caso específico: como dimensionar uma prova ideal para a cidade de Sertãozinho.

De forma econômico-político-geográfica, Sertãozinho pode ser descrita com base em <http://www.sertaozinho.com.br/localizacao.shtml> (27/03/2005), conforme texto a seguir

A cidade de Sertãozinho, com uma população de aproximadamente 100 mil habitantes, está localizada a nordeste do Estado de São Paulo, na zona fisiográfica de Ribeirão Preto, distanciando em linha reta, 305 Km da capital do estado de SP - a cidade de São Paulo.

Sertãozinho limita-se ao norte com Jardinópolis e Pontal; a oeste com Jaboticabal e Pitangueiras; ao sul com Barrinha e Dumont e a leste com Ribeirão Preto (distante a 21 km).

É uma das cidades mais importantes da região, sendo a primeira produtora de açúcar e álcool do Estado de São Paulo contando também com um grande parque industrial.

Mapa Estilizado do Município de Sertãozinho
(fonte: www.sertaozinho.com.br)

O distrito de Sertãozinho foi criado pela Lei Provincial de No. 31, de 10 de março de 1885.

Em 5 de dezembro de 1896, através da Lei Estadual No. 463, tornou-se município, tendo sede na povoação Aparecida de Sertãozinho (território desmembrado do Município de Ribeirão Preto). Nesta época, contava com uma área de 1.071 Km².

A Lei estadual No. 1038, de 10 de dezembro de 1906, elevou a sede do município a categoria de cidade.

O município de Sertãozinho teve sua área diminuída por várias vezes por ocasião das divisões administrativas.

Inicialmente houve o desmembramento do distrito de Pontal, que tornou-se município (1935). Durante o governo da ditadura, Sertãozinho aos poucos foi cedendo sua área a municípios vizinhos, sem atender a nenhuma razão histórica nem divisas naturais. Em 1938, ocorreu o desmembramento do município de Pradópolis e, em 1953, com a criação do município de Barrinha, restringiu-se mais ainda sua área.

Região de RIBEIRÃO PRETO (16 cidades pertencem a esta região)

Barrinha, Brodósqui, Cravinhos, Dumont, Guataporã, Jardinópolis, Luís Antônio, Pontal, Pradópolis, Ribeirão Preto, Santa Rita do Passa Quatro, Santa Rosa de Viterbo, São Simão, Serra Azul, Serrana, Sertãozinho

Posição geográfica da cidade de Sertãozinho

21° 08' 33" latitude sul (-21,13778); 47° 59' 09" longitude oeste (-47,99028) sendo a sede do município como referência.

Altitude: 579 m

Área Territorial Total: 403 Km²

Ante todas essas informações, cabe agora refletir como elas se correlacionam com as habilidades/proficiências dos alunos da rede SESI no município.

Faz sentido assumir uma distribuição de habilidades acima da média do Estado de São Paulo ante os indicadores geo-político-demográficos. Maiores informações sobre como quantificar (modificar) nossa incerteza acerca das habilidades dos examinandos por hora estão ausentes. Ou seja, não possuímos uma verossimilhança (informação) capaz de melhorar nosso conhecimento acerca dos desempenhos dos alunos da rede SESI com base nos indicadores geo-político-demográficos disponíveis.

Em contrapartida, a informação sobre o desempenho da 8ª série em Matemática no ano de 2002 para o sistema SESI de São Paulo, passa a ser eventualmente interessante. Se considerarmos que as alterações nos indicadores geo-político-demográficos-educacionais, foram pouco relevantes no período, podemos considerar que as habilidades dos examinandos atuais (2005) sejam igualmente distribuídas com referência ao ano de 2002.

Deste modo a Prova Ideal, em 2005, em Matemática, para a 8ª série, poderia ser escolhida com base no desempenho do município em 2002, cujas médias e variâncias foram 0,683152598244986 e 0,616336891615473, respectivamente. Simplificadamente podemos assumir $\theta \sim N(0.683152598244986, 0.616336891615473)$ como a priori para os examinandos de 2005.

Considerando essas verossimilhanças e prioris (intrinsecamente subjetivas), temos

Na avaliação do SESI em Matemática, realizada em 2002, foram obtidos os seguintes dados:

Número de Alunos que realizaram a prova de Matemática: 111

Nota Média dos Alunos*: 0,683152598244986

Variância das Notas dos Alunos*: 0,616336891615473

* = Escala de Notas elaborada através da Teoria da Resposta ao Item para todo o sistema SESI do Estado de São Paulo (13.717 alunos avaliados) com escala de habilidades fixada em média (SP) = 0,00783367172237199 e variância (SP) = 0,840525544305572.

Considerando uma simulação de $\theta \sim N(0.683152598244986, 0.616336891615473)$, referente ao atual conhecimento das habilidades dos examinandos de Sertãozinho, os Riscos de Bayes acerca dos itens de uma prova são dados pela tabela a seguir; para a definição da Prova Ótima, para Sertãozinho, basta escolher os k itens de menor Risco de Bayes, ordenadamente.

Ordem	Rótulo do Item j	a	b	c	Esperança de Acerto	Risco de Bayes
01	0024	1,286593	0,961594	0,056341	45,89%	0,5248909
02	0012	1,737807	1,463528	0,105928	34,41%	0,5265633
03	0081	1,575579	1,061785	0,157240	48,39%	0,5274724
04	0089	1,439171	0,806498	0,170633	55,60%	0,5318622
05	0016	1,191994	0,545207	0,103734	58,29%	0,5396745
06	0017	1,574178	1,088921	0,221444	51,63%	0,5414432
07	0084	1,294290	0,978026	0,141588	50,39%	0,5417187
08	0057	1,368259	1,209117	0,137943	44,62%	0,5421008
09	0079	1,438180	1,351768	0,155267	42,31%	0,5467374
10	0075	1,244086	1,424430	0,090181	37,84%	0,5486103
11	0009	1,131747	0,693016	0,188866	59,21%	0,5571320
12	0063	1,100874	0,676354	0,180141	59,04%	0,5585242
13	0080	0,942468	1,376307	0,023080	37,35%	0,5593035
14	0083	1,689588	1,405610	0,288540	49,10%	0,5640810
15	0072	1,090451	0,727405	0,235029	60,92%	0,5659546
16	0054	1,387393	1,863333	0,122715	30,49%	0,5664894
17	0032	0,931014	0,741297	0,125737	55,23%	0,5669174
18	0010	1,031067	1,794677	0,063795	31,40%	0,5669409
19	0028	1,041062	1,217459	0,161567	48,03%	0,5681445
20	0013	1,080700	1,290500	0,173604	47,25%	0,5685741
21	0023	1,195212	1,671080	0,148106	37,65%	0,5693494
22	0045	1,246359	1,253080	0,265066	52,66%	0,5703145
23	0052	1,588452	1,846751	0,174898	33,20%	0,5707149
24	0076	0,918962	-0,260358	0,137519	72,90%	0,5712083
25	0069	0,871425	0,795611	0,156571	55,96%	0,5754226
26	0019	1,341619	1,560848	0,256837	46,10%	0,5754617
27	0041	1,142205	1,628624	0,183745	41,53%	0,5754630
28	0040	0,854589	-0,185284	0,140492	71,00%	0,5756480
29	0046	1,004794	1,790861	0,109781	35,05%	0,5758016
30	0064	0,882782	0,434281	0,188767	63,40%	0,5759904
31	0030	0,919421	1,046807	0,173345	52,50%	0,5763073
32	0048	1,140378	1,696178	0,175265	39,85%	0,5763526
33	0050	0,929479	1,454480	0,138288	43,53%	0,5768581
34	0085	0,810795	0,564407	0,130582	58,37%	0,5771833
35	0051	0,971438	1,651490	0,137687	39,87%	0,5776972
36	0034	0,892463	-0,982224	0,064935	80,72%	0,5783301
37	0053	0,871619	0,996733	0,184200	54,16%	0,5788618
38	0087	0,979064	1,287396	0,221658	51,20%	0,5798923
39	0086	1,225780	2,023583	0,142983	31,32%	0,5799275
40	0061	1,796632	1,912899	0,219107	34,41%	0,5800343
41	0077	0,734795	0,445803	0,080037	57,73%	0,5805608
42	0038	0,984062	-1,319818	0,059346	86,38%	0,5808742
43	0056	1,198665	1,681138	0,243088	44,47%	0,5816742
44	0059	0,986934	1,818890	0,144966	37,53%	0,5818973
45	0068	0,789092	-0,614838	0,100306	74,79%	0,5822874
46	0055	0,834279	-0,870724	0,124220	79,55%	0,5833130
47	0018	0,712116	-0,072507	0,093768	65,81%	0,5839330
48	0031	0,727212	1,279533	0,085123	45,11%	0,5839862
49	0026	1,066476	1,827624	0,203087	40,72%	0,5848453
50	0044	0,709947	1,642380	0,058419	38,41%	0,5857406
51	0060	0,991964	1,360292	0,287482	54,15%	0,5863785
52	0082	0,728038	0,212316	0,191512	65,96%	0,5876934
53	0074	1,165454	1,899933	0,235066	41,00%	0,5878350
54	0014	0,728578	1,094939	0,163796	52,45%	0,5881707
55	0035	0,707262	-0,660946	0,126080	74,52%	0,5888960
56	0065	0,810860	1,176142	0,236469	54,91%	0,5892673
57	0070	0,772046	1,784901	0,144030	41,15%	0,5906158
58	0021	0,715881	-0,900203	0,149559	78,17%	0,5910520
59	0007	0,617861	0,761681	0,105489	54,22%	0,5912707
60	0067	0,710845	-1,215786	0,103602	80,31%	0,5919437
61	0003	0,656179	0,292201	0,181445	63,99%	0,5921629
62	0022	0,781366	1,705218	0,199823	45,88%	0,5934498
63	0015	0,874349	1,250634	0,348445	60,22%	0,5941071
64	0049	0,588143	0,173243	0,150428	63,57%	0,5956964
65	0078	0,597693	1,761395	0,097116	41,37%	0,5965084
66	0027	0,771562	2,133933	0,202337	41,07%	0,5986641
67	0029	0,660295	-1,781839	0,141006	84,86%	0,5992106
68	0037	0,504869	-0,409659	0,104636	66,85%	0,6000254
69	0073	0,538325	0,053542	0,192716	66,11%	0,6005966
70	0043	0,522398	0,609031	0,171718	59,34%	0,6007019
71	0001	0,493367	0,264156	0,129617	60,94%	0,6012761
72	0042	0,557044	1,631573	0,173526	48,49%	0,6013807
73	0039	0,643799	-2,232272	0,078630	86,83%	0,6014441
74	0020	0,673200	0,139471	0,429941	76,38%	0,6015685
75	0058	0,648976	2,511292	0,169540	37,24%	0,6025232
76	0008	0,735503	2,358819	0,238117	42,14%	0,6030847
77	0088	0,720697	-2,499473	0,118976	91,02%	0,6031175
78	0033	0,522896	-2,020800	0,104420	81,82%	0,6040120
79	0002	0,494281	-1,397275	0,152413	77,18%	0,6040546
80	0090	0,527697	-1,329215	0,261209	80,50%	0,6048921
81	0066	0,958506	2,855891	0,182047	28,93%	0,6049081
82	0004	0,458941	1,291023	0,186553	53,91%	0,6057443
83	0025	0,427965	-0,124878	0,166460	65,27%	0,6060134
84	0036	0,459502	-2,185127	0,068518	79,79%	0,6060930
85	0005	0,392948	0,578515	0,185217	60,11%	0,6082841
86	0006	0,279520	-1,164538	0,199908	70,00%	0,6130345
87	0071	0,150154	-0,848973	0,272241	67,83%	0,6166223
88	0011	0,109092	9,443419	0,080025	33,61%	0,6171811
89	0047	0,047427	17,587870	0,013351	32,02%	0,6177286
90	0062	0,046751	13,823580	0,021471	36,46%	0,6177405

Tabela 8: *Ranking* dos itens SESI 2002 ordenados de acordo com o Risco de Bayes considerando a priori $\theta \sim N(0.683152598244986, 0.616336891615473)$, obtida via simulação.

Este tópico visa estabelecer uma análise de caso comparativa entre as provas T, U e V, aplicadas para os alunos da 8ª série do Ensino Fundamental na rede de ensino SESI em 2002. A distribuição a priori das habilidades dos examinandos considerada para a melhor escolha dos itens foi $\theta \sim N(0, 1)$.

Desse modo os itens foram ordenados através dos seus *ranking's* medidos pelo Risco de Bayes; a melhor prova, por conseguinte, é aquela que contém os melhores itens, ou seja, os primeiros itens ordenados através dos seus *ranking's*.

Nesse contexto, a coluna "*Ranking*" nas tabelas 9 e 10, a seguir, indicam a posição (*ranking*) do item j para a escolha dos melhores itens com uma priori $\theta \sim N(0, 1)$; a coluna Rótulo do item j indica o "nome" do item j ; prosseguindo com a descrição, nas tabelas 9 e 10, as colunas a , b e c do grupo "Parâmetros dos Itens" indicam os valores dos parâmetros dos itens j na Teoria da Resposta ao Item, no modelo logístico de 3 parâmetros; as colunas "Probabilidade de Acerto" e "Risco de Bayes" indicam esses valores para a priori $\theta \sim N(0, 1)$; adicionalmente, as colunas T, U e V no grupo "Indicador de Presença na Prova" e "*Ranking* do Item Presente na Prova" assinalam os itens j presentes nas provas T, U e V e seus respectivos *Ranking's* no conjunto total de itens; a coluna "Indicador de Item Comum" indica os itens comuns a 2 ou 3 provas (o valor 2 indica a presença em duas provas e o valor 3, em três provas); e, finalmente, a coluna "*Ranking* dos Itens Comuns" indica a posição (*ranking*) dos itens presentes a duas ou mais provas.

Nessa análise de caso, de comparação das provas via *ranking*, ilustrada pela tabela 9, verificamos que a prova V sobressai-se em relação às provas T e U: ela possui um número relevantemente maior de itens com bom *ranking* e poucos itens de *ranking* ruim em relação às provas T e U. O *ranking* médio na prova V é 41,48 contra os *ranking's* médios de 46,80 e 46,38 das provas U e V, respectivamente. O *ranking* médio da prova V sobressai-se, inclusive, diante dos itens comuns entre duas e/ou três provas, cujo *ranking* médio é 45,75. Nesse contexto torna-se relevante assinalar que, do ponto de vista teórico, dispondo do mesmo conjunto de itens, os "itens comuns" a duas e/ou três provas deveriam ser, necessariamente, os melhores itens para o universo-alvo, fato que não se verifica nesse estudo de caso. Ou seja, com essa análise de caso, dispomos de indicadores sólidos de que a escolha dos itens de uma prova pode ser melhorada com a teoria proposta no transcórre deste texto.

Para ilustrar essa potencialidade foi construída a tabela 10: essa tabela foi construída empiricamente, respeitando-se os princípios teóricos-práticos apresentados. Comparando-se as tabelas 9 e 10 verificamos uma melhora imediata no *ranking* médio de todas as provas T', U' e V' (tabela 10) em relação às provas T, U e V originais (tabela 9); o *ranking* médio dos itens comuns a 2 e/ou 3 provas foi melhorado significativamente: ele (*ranking* médio) foi reduzido de 45,75 nas provas T, U e V originais para 10,50 nas provas T', U' e V' propostas. Ressalte-se que todos esses indicadores das provas foram melhorados com base no mesmo subconjunto de itens, com a mesma quantidade de itens comuns a 2 e/ou 3 provas e, para a mesma população.

Ranking	Rótulo do Item j	Parâmetros do Item j			Probabilidade de Acerto $E[X_j=1]$	Risco de Bayes	Indicador de presença do item / na Prova			Ranking do Item / Presente na Prova			Indicador de Item Comum	Ranking dos Itens Comuns
		a	b	c			T	U	V	T	U	V		
1	0024	1,2866	0,9616	0,0563	0,3195	0,8342	1			1				
2	0016	1,1920	0,5452	0,1037	0,4405	0,8446	1			2				
3	0089	1,4392	0,8065	0,1706	0,4201	0,8597			1			3		
4	0081	1,5756	1,0618	0,1572	0,3555	0,8695			1			4		
5	0084	1,2943	0,9780	0,1416	0,3787	0,8745			1			5		
6	0038	0,9841	-1,3198	0,0593	0,7633	0,8783	1	1	1	6	6	6	2	6
7	0034	0,8925	-0,9822	0,0649	0,7005	0,8836	1	1	1	7	7	7	2	7
8	0076	0,9190	-0,2604	0,1375	0,6124	0,8847			1			8		
9	0080	0,9425	1,3763	0,0231	0,2670	0,8867			1			9		
10	0012	1,7378	1,4635	0,1059	0,2409	0,8868	1			10				
11	0057	1,3683	1,2091	0,1379	0,3312	0,8869		1			11			
12	0009	1,1317	0,6930	0,1889	0,4711	0,8885	1		1	12		12	1	12
13	0063	1,1009	0,6764	0,1801	0,4703	0,8895		1			13			
14	0075	1,2441	1,4244	0,0902	0,2718	0,8926			1			14		
15	0017	1,5742	1,0889	0,2214	0,4015	0,8976	1			15				
16	0040	0,8546	-0,1853	0,1405	0,5993	0,8979	1	1	1	16	16	16	2	16
17	0032	0,9310	0,7413	0,1257	0,4380	0,8988	1	1	1	17	17	17	2	17
18	0055	0,8343	-0,8707	0,1242	0,6962	0,9019		1			18			
19	0068	0,7891	-0,6148	0,1003	0,6446	0,9025		1			19			
20	0079	1,4382	1,3518	0,1553	0,3176	0,9041			1			20		
21	0072	1,0905	0,7274	0,2350	0,4978	0,9070			1			21		
22	0064	0,8828	0,4343	0,1888	0,5287	0,9119		1			22			
23	0085	0,8108	0,5644	0,1306	0,4787	0,9127			1			23		
24	0077	0,7348	0,4458	0,0800	0,4727	0,9137			1			24		
25	0069	0,8714	0,7956	0,1566	0,4555	0,9160		1			25			
26	0018	0,7121	-0,0725	0,0938	0,5567	0,9162	1			26				
27	0028	1,0411	1,2175	0,1616	0,3798	0,9170	1			27				
28	0010	1,0311	1,7947	0,0638	0,2279	0,9181	1			28				
29	0013	1,0807	1,2905	0,1736	0,3743	0,9210	1		1			29	1	29
30	0030	0,9194	1,0468	0,1733	0,4259	0,9226	1			30				
31	0035	0,7073	-0,6609	0,1261	0,6531	0,9236	1	1	1	31	31	31	2	31
32	0067	0,7108	-1,2158	0,1036	0,7186	0,9255		1			32			
33	0021	0,7159	-0,9002	0,1496	0,6942	0,9260	1			33				
34	0053	0,8716	0,9967	0,1842	0,4460	0,9286		1			34			
35	0050	0,9295	1,4545	0,1383	0,3448	0,9295		1			35			
36	0031	0,7272	1,2795	0,0851	0,3614	0,9314	1			36				
37	0082	0,7280	0,2123	0,1915	0,5676	0,9324			1			37		
38	0023	1,1952	1,6711	0,1481	0,2913	0,9329	1	1	1	38	38	38	2	38
39	0045	1,2464	1,2531	0,2651	0,4336	0,9334		1			39			
40	0046	1,0048	1,7909	0,1098	0,2699	0,9342		1			40			
41	0044	0,7099	1,6424	0,0584	0,3016	0,9359		1	1		41	41	1	41
42	0051	0,9714	1,6515	0,1377	0,3144	0,9363		1			42			
43	0054	1,3874	1,8633	0,1227	0,2297	0,9376		1			43			
44	0087	0,9791	1,2874	0,2217	0,4227	0,9382			1			44		
45	0041	1,1422	1,6286	0,1837	0,3317	0,9402		1			45			
46	0014	0,7286	1,0949	0,1638	0,4384	0,9414	1			46				
47	0007	0,6179	0,7617	0,1055	0,4577	0,9415	1			47			1	47
48	0083	1,6896	1,4056	0,2885	0,4044	0,9419			1			48		
49	0048	1,1404	1,6962	0,1753	0,3174	0,9423		1			49			
50	0003	0,6562	0,2922	0,1814	0,5546	0,9424	1			50				
51	0029	0,6603	-1,7818	0,1410	0,7814	0,9436	1			51			1	51
52	0059	0,9869	1,8189	0,1450	0,2984	0,9442		1			52			
53	0065	0,8109	1,1761	0,2365	0,4678	0,9475		1			53			
54	0039	0,6438	-2,2323	0,0786	0,8070	0,9477	1	1	1	54	54	54	2	54
55	0049	0,5881	0,1732	0,1504	0,5551	0,9479		1			55			
56	0019	1,3416	1,5608	0,2568	0,3818	0,9481	1			56				
57	0088	0,7207	-2,4995	0,1190	0,8569	0,9482			1			57		
58	0070	0,7720	1,7849	0,1440	0,3383	0,9512		1			58			
59	0052	1,5885	1,8468	0,1749	0,2632	0,9516		1	1		59	59	1	59
60	0060	0,9920	1,3603	0,2875	0,4618	0,9518		1			60			
61	0086	1,2258	2,0236	0,1430	0,2467	0,9522			1			61		
62	0056	1,1987	1,6811	0,2431	0,3695	0,9535		1			62			
63	0026	1,0665	1,8276	0,2031	0,3344	0,9547	1			63				
64	0037	0,5049	-0,4097	0,1046	0,5962	0,9548	1	1	1	64	64	64	2	64
65	0078	0,5977	1,7614	0,0971	0,3439	0,9566			1			65		
66	0022	0,7814	1,7052	0,1998	0,3875	0,9575	1		1	66			1	66
67	0073	0,5383	0,0535	0,1927	0,5902	0,9588			1			67		
68	0033	0,5229	-2,0208	0,1044	0,7586	0,9598	1	1	1	68	68	68	2	68
69	0001	0,4934	0,2642	0,1296	0,5373	0,9601			1		69			
70	0015	0,8743	1,2506	0,3484	0,5306	0,9610	1			70				
71	0043	0,5224	0,6090	0,1717	0,5245	0,9614		1	1		71	71	1	71
72	0002	0,4943	-1,3973	0,1524	0,7104	0,9627	1			72				
73	0074	1,1655	1,8999	0,2351	0,3441	0,9634			1			73		
74	0090	0,5277	-1,3292	0,2612	0,7479	0,9654			1			74		
75	0036	0,4595	-2,1851	0,0685	0,7417	0,9659	1	1	1	75	75	75	2	75
76	0020	0,6732	0,1395	0,4299	0,7029	0,9661	1			76				
77	0061	1,7966	1,9129	0,2191	0,2872	0,9661		1	1		77	77	1	77
78	0042	0,5570	1,6316	0,1735	0,4207	0,9673		1			78			
79	0027	0,7716	2,1339	0,2023	0,3515	0,9695	1			79				
80	0025	0,4280	-0,1249	0,1665	0,5937	0,9712	1			80				
81	0004	0,4589	1,2910	0,1866	0,4813	0,9750	1			81				
82	0058	0,6490	2,5113	0,1695	0,3205	0,9758		1			82			
83	0008	0,7355	2,3588	0,2381	0,3700	0,9782			1		83			
84	0005	0,3929	0,5785	0,1852	0,5482	0,9788	1				84			
85	0066	0,9585	2,8559	0,1820	0,2512	0,9858		1			85			
86	0006	0,2795	-1,1645	0,1999	0,6634	0,9888	1		1	86		86	1	86
87	0071	0,1502	-0,8490	0,2722	0,6594	0,9984			1			87		
88	0011	0,1091	9,4434	0,0800	0,3227	1,0000	1			88				
89	0047	0,0474	17,5879	0,0134	0,3120	1,0013		1			89			
90	0062	0,0468	13,8236	0,0215	0,3579	1,0014		1			90			
SOMA							40	40	40	46,80	46,38	41,48	30	45,75
MÉDIA														

Tabela 9: Análise de Caso entre Provas Realizadas. Comparação dos *ranking's* dos itens das provas T, U e V aplicadas no sistema SESI em 2002.

Ranking	Rótulo do Item j	Parâmetros do Item j			Probabilidade de Acerto $E[X_j=1]$	Risco de Bayes	Indicador de presença do item / na Prova			Ranking do Item j / Presente na Prova			Indicador de Item Comum	Ranking dos Itens Comuns
		a	b	c			T'	U'	V'	T'	U'	V'		
1	0024	1,2866	0,9616	0,0563	0,3195	0,8342	1	1	1	1	1	1	2	1
2	0016	1,1920	0,5452	0,1037	0,4405	0,8446	1	1	1	2	2	2	2	2
3	0089	1,4392	0,8065	0,1706	0,4201	0,8597	1	1	1	3	3	3	2	3
4	0081	1,5756	1,0618	0,1572	0,3555	0,8695	1	1	1	4	4	4	2	4
5	0084	1,2943	0,9780	0,1416	0,3787	0,8745	1	1	1	5	5	5	2	5
6	0038	0,9841	-1,3198	0,0593	0,7633	0,8783	1	1	1	6	6	6	2	6
7	0034	0,8925	-0,9822	0,0649	0,7005	0,8836	1	1	1	7	7	7	2	7
8	0076	0,9190	-0,2604	0,1375	0,6124	0,8847	1	1	1	8	8	8	2	8
9	0080	0,9425	1,3763	0,0231	0,2670	0,8867	1	1	1	9	9	9	2	9
10	0012	1,7378	1,4635	0,1059	0,2409	0,8868	1	1	1	10	10	10	2	10
11	0057	1,3683	1,2091	0,1379	0,3312	0,8869	1	1	1	11	11	11	1	11
12	0009	1,1317	0,6930	0,1889	0,4711	0,8885	1	1	1	12	12	12	1	12
13	0063	1,1009	0,6764	0,1801	0,4703	0,8895	1	1	1	13	13	13	1	13
14	0075	1,2441	1,4244	0,0902	0,2718	0,8926	1	1	1	14	14	14	1	14
15	0017	1,5742	1,0889	0,2214	0,4015	0,8976	1	1	1	15	15	15	1	15
16	0040	0,8546	-0,1853	0,1405	0,5993	0,8979	1	1	1	16	16	16	1	16
17	0032	0,9310	0,7413	0,1257	0,4380	0,8988	1	1	1	17	17	17	1	17
18	0055	0,8343	-0,8707	0,1242	0,6962	0,9019	1	1	1	18	18	18	1	18
19	0068	0,7891	-0,8148	0,1003	0,6446	0,9025	1	1	1	19	19	19	1	19
20	0079	1,4382	1,3518	0,1553	0,3176	0,9041	1	1	1	20	20	20	1	20
21	0072	1,0905	0,7274	0,2350	0,4978	0,9070	1	1	1	21	21	21		
22	0064	0,8828	0,4343	0,1888	0,5287	0,9119	1	1	1	22	22	22		
23	0085	0,8108	0,5644	0,1306	0,4787	0,9127	1	1	1	23	23	23		
24	0077	0,7348	0,4458	0,0800	0,4727	0,9137	1	1	1	24	24	24		
25	0069	0,8714	0,7956	0,1566	0,4555	0,9160	1	1	1	25	25	25		
26	0018	0,7121	-0,0725	0,0938	0,5567	0,9162	1	1	1	26	26	26		
27	0028	1,0411	1,2175	0,1616	0,3798	0,9170	1	1	1	27	27	27		
28	0010	1,0311	1,7947	0,0638	0,2279	0,9181	1	1	1	28	28	28		
29	0013	1,0807	1,2905	0,1736	0,3743	0,9210	1	1	1	29	29	29		
30	0030	0,9194	1,0468	0,1733	0,4259	0,9226	1	1	1	30	30	30		
31	0035	0,7073	-0,8609	0,1261	0,6531	0,9236	1	1	1	31	31	31		
32	0067	0,7108	-1,2158	0,1036	0,7186	0,9255	1	1	1	32	32	32		
33	0021	0,7159	-0,9002	0,1496	0,6942	0,9260	1	1	1	33	33	33		
34	0053	0,8716	0,9967	0,1842	0,4460	0,9286	1	1	1	34	34	34		
35	0050	0,9295	1,4545	0,1383	0,3448	0,9295	1	1	1	35	35	35		
36	0031	0,7272	1,2795	0,0851	0,3614	0,9314	1	1	1	36	36	36		
37	0082	0,7280	0,2123	0,1915	0,5676	0,9324	1	1	1	37	37	37		
38	0023	1,1952	1,6711	0,1481	0,2913	0,9329	1	1	1	38	38	38		
39	0045	1,2464	1,2531	0,2651	0,4336	0,9334	1	1	1	39	39	39		
40	0046	1,0048	1,7909	0,1098	0,2699	0,9342	1	1	1	40	40	40		
41	0044	0,7099	1,6424	0,0584	0,3016	0,9359	1	1	1	41	41	41		
42	0051	0,9714	1,6515	0,1377	0,3144	0,9363	1	1	1	42	42	42		
43	0054	1,3874	1,8633	0,1227	0,2297	0,9376	1	1	1	43	43	43		
44	0087	0,9791	1,2874	0,2217	0,4227	0,9382	1	1	1	44	44	44		
45	0041	1,1422	1,6286	0,1837	0,3317	0,9402	1	1	1	45	45	45		
46	0014	0,7286	1,0949	0,1638	0,4384	0,9414	1	1	1	46	46	46		
47	0007	0,6179	0,7617	0,1055	0,4577	0,9415	1	1	1	47	47	47		
48	0083	1,6896	1,4056	0,2885	0,4044	0,9419	1	1	1	48	48	48		
49	0048	1,1404	1,6962	0,1753	0,3174	0,9423	1	1	1	49	49	49		
50	0003	0,6562	0,2922	0,1814	0,5546	0,9424	1	1	1	50	50	50		
51	0029	0,6603	-1,7818	0,1410	0,7814	0,9436	1	1	1	51	51	51		
52	0059	0,9869	1,8189	0,1450	0,2984	0,9442	1	1	1	52	52	52		
53	0065	0,8109	1,1761	0,2365	0,4678	0,9475	1	1	1	53	53	53		
54	0039	0,6438	-2,2323	0,0786	0,8070	0,9477	1	1	1	54	54	54		
55	0049	0,5881	0,1732	0,1504	0,5551	0,9479	1	1	1	55	55	55		
56	0019	1,3416	1,5608	0,2568	0,3818	0,9481	1	1	1	56	56	56		
57	0088	0,7207	-2,4995	0,1190	0,8569	0,9482	1	1	1	57	57	57		
58	0070	0,7720	1,7849	0,1440	0,3383	0,9512	1	1	1	58	58	58		
59	0052	1,5885	1,8468	0,1749	0,2632	0,9516	1	1	1	59	59	59		
60	0060	0,9920	1,3603	0,2875	0,4618	0,9518	1	1	1	60	60	60		
61	0086	1,2258	2,0236	0,1430	0,2467	0,9522	1	1	1	61	61	61		
62	0056	1,1987	1,6811	0,2431	0,3695	0,9535	1	1	1	62	62	62		
63	0026	1,0665	1,8276	0,2031	0,3344	0,9547	1	1	1	63	63	63		
64	0037	0,5049	-0,4097	0,1046	0,5962	0,9548	1	1	1	64	64	64		
65	0078	0,5977	1,7614	0,0971	0,3439	0,9566	1	1	1	65	65	65		
66	0022	0,7814	1,7052	0,1998	0,3875	0,9575	1	1	1	66	66	66		
67	0073	0,5383	0,0535	0,1927	0,5902	0,9588	1	1	1	67	67	67		
68	0033	0,5229	-2,0208	0,1044	0,7586	0,9598	1	1	1	68	68	68		
69	0001	0,4934	0,2642	0,1296	0,5373	0,9601	1	1	1	69	69	69		
70	0015	0,8743	1,2506	0,3484	0,5306	0,9610	1	1	1	70	70	70		
71	0043	0,5224	0,6090	0,1717	0,5245	0,9614	1	1	1	71	71	71		
72	0002	0,4943	-1,3973	0,1524	0,7104	0,9627	1	1	1	72	72	72		
73	0074	1,1655	1,8999	0,2351	0,3441	0,9634	1	1	1	73	73	73		
74	0090	0,5277	-1,3292	0,2612	0,7479	0,9654	1	1	1	74	74	74		
75	0036	0,4595	-2,1851	0,0685	0,7417	0,9659	1	1	1	75	75	75		
76	0020	0,6732	0,1395	0,4299	0,7029	0,9661	1	1	1	76	76	76		
77	0061	1,7966	1,9129	0,2191	0,2872	0,9661	1	1	1	77	77	77		
78	0042	0,5570	1,6316	0,1735	0,4207	0,9673	1	1	1	78	78	78		
79	0027	0,7716	2,1339	0,2023	0,3515	0,9695	1	1	1	79	79	79		
80	0025	0,4280	-0,1249	0,1665	0,5937	0,9712	1	1	1	80	80	80		
81	0004	0,4589	1,2910	0,1866	0,4813	0,9750	1	1	1	81	81	81		
82	0058	0,6490	2,5113	0,1695	0,3205	0,9758	1	1	1	82	82	82		
83	0008	0,7355	2,3588	0,2381	0,3700	0,9782	1	1	1	83	83	83		
84	0005	0,3929	0,5785	0,1852	0,5482	0,9788	1	1	1	84	84	84		
85	0066	0,9585	2,8559	0,1820	0,2512	0,9858	1	1	1	85	85	85		
86	0006	0,2795	-1,1645	0,1999	0,6634	0,9888	1	1	1	86	86	86		
87	0071	0,1502	-0,8490	0,2722	0,6594	0,9984	1	1	1	87	87	87		
88	0011	0,1091	9,4434	0,0800	0,3227	1,0000	1	1	1	88	88	88		
89	0047	0,0474	17,5879	0,0134	0,3120	1,0013	1	1	1	89	89	89		
90	0062	0,0468	13,8236	0,0215	0,3579	1,0014	1	1	1	90	90	90		
SOMA							40	40	40				30	
MÉDIA										36,60	36,25	36,15		10,50

Tabela 10: Análise de Caso entre Provas Realizadas. Proposta melhorada "empiricamente" para a escolha dos itens das provas T', U' e V' com base na população do sistema SESI de 2002.

9) Conclusão

No texto apresentado vimos que quando consideramos a Variância como função de Incerteza U , de natureza quadrática, o experimento que maximiza a Informação $I\{X; P, U\} = U(P) - E[U(P_x)]$

é dada por $J^* = \arg \max_J I\{X_J; \xi, V\} = V(\theta) - E[V(\theta | X_J)] = V(E[\theta | X_J])$, apresentado como

Experimento de Máxima Variância de Bayes. O Experimento de Máxima Variância de Bayes tem a propriedade de ser o experimento mais sensível aos dados amostrais.

Adicionalmente estabelecemos um paralelo entre o Experimento de Máxima Variância de Bayes e o Experimento Ótimo à Luz da Teoria da Decisão, quando a função de perda $L(d, \theta, X_{d_j}) = (\theta - E[\theta | X_{d_j}])^2$ é de natureza quadrática. Dessa forma, quando a função de Incerteza U (Experimento de Bayes) e a função de Perda L (Teoria da Decisão) são consistentemente definidas, a escolha dos experimentos é equivalente nas duas abordagens, resultando na escolha dos mesmos experimentos devido à propriedade $V(\theta) = V(E[\theta | X]) + E[V(\theta | X)]$.

Os casos da escolha de k Experimentos, sem repetição, assim como a escolha sequencial de k experimentos, dentre n experimentos disponíveis, também foram abordados e o resultado é que devemos utilizar uma análise pré-posteriori nesses casos. Quando desejamos obter k experimentos dentre n experimentos disponíveis de uma única vez, sem repetição de experimentos, o resultado é que podemos selecioná-los mediante *ranking* do Risco de Bayes; quando desejamos obter k experimentos dentre n experimentos disponíveis sequencialmente, sem repetição, o resultado é que devemos utilizar a posteriori $\theta | x_{1,...,t-1}$ como a "priori" na escolha de X_t , propriedade essa, conhecida como equivalência entre as formas normal e extensiva da Teoria da Decisão.

Estabelecida a base teórica apresentamos a Teoria da Resposta ao Item como elo para a escolha dos melhores itens de uma prova, onde buscamos maximizar a Informação sobre a Habilidade θ dos examinandos. A escolha dos itens de uma prova pode ser vista como a escolha dos melhores Experimentos de Bayes de forma que, aplicando a base teórica, obtemos a escolha do(s) melhor(es) item(ns) de uma prova com embasamento estatístico. Dadas as propriedades da Teoria da Decisão/Experimento de Bayes, a escolha dos itens por esse método terá propriedades ótimas: qualquer outro método, por exemplo empírico, será menos Informativo a respeito da habilidade θ dos examinandos; desse modo, considerando que as provas atuais são compostas de itens escolhidos empiricamente por especialistas da área, é possível melhorar essa escolha com método proposto, auferindo-se uma precisão (Informação) equivalente com menos itens na prova. Por conseguinte é possível uma redução de custos e prazos associados à realização de uma prova, mediante a escolha dos itens pelo método proposto.

Bibliografia

- Andrade, D., Valle, R. (1998). Introdução à Teoria da Resposta ao Item. *Estudos em Avaliação Educacional*. São Paulo: Fundação Carlos Chagas, **18**, 13-32.
- Angoff, W. (1971). Scales, Norms, and Equivalent Scores. In R. L. Thorndike (Ed.), *Educational Measurement (2nd. ed.)*, Washington, D.C.: American Council on Education.
- Basu, D. (1975). Statistical information and likelihood (with discussion). *Sankhya A* 37 1-71.
- Basu, D., Pereira, C. (1990). Blackwell Sufficiency and Bernoulli Experiments. *Rebrap*, **4**, 137-45.
- DeGroot, M. (1962). Uncertainty, Information, and Sequential Experiments. *Ann. Math. Statistics* **33**, 404-19.
- DeGroot, M. (1970). Optimal Statistical Decisions. *McGraw Hill*, New York.
- Gulliksen, H. (1950). Theory of Mental Tests. *New York: Wiley*.
- Klein, R. (2003). Utilização da Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). *Ensaio: Avaliação e Políticas Públicas em Educação*. Rio de Janeiro, **11**, **40**, 283-96.
- Mislevy, J., Bock, D (1990). BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models. Chicago: Scientific Software Inc.
- Lord, F. (1977). Practical Applications of Item Response Theory to Practical Testing Problems. *Hillsdale, New Jersey: Lawrence Erlbaum*.
- Lord, F., Novick, M. (1968). Statistical Theories of Mental Test Scores. *Reading Mass.: Addison-Wesley*.
- Valle, R. (2000). Teoria da Resposta ao Item. *Estudos em Avaliação Educacional*. São Paulo: Fundação Carlos Chagas, **21**, 7-91.
- Valle, R. (2001). Construção e Interpretação de Escalas de Conhecimento: um Estudo de Caso. *Estudos em Avaliação Educacional*. São Paulo: Fundação Carlos Chagas, **23**, 71-92.