

Aspectos Inferenciais em Experimentos de Bernoulli com Erros e Classificações Repetidas

Joab de Oliveira Lima
(joab@de.ufpb.br)

Tese submetida ao Programa de Pós-graduação em Estatística da
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Em 09/12/2009

em cumprimento às exigências para a obtenção do grau de
DOUTOR EM ESTATÍSTICA

Área de Concentração: Estatística

Orientador: Prof. Roberto da Costa Quinino
(roberto@est.ufmg.br)

À minha esposa amada Jucilene e aos meus
filhos Letícia e Nicolás.

Todas as vezes que você perdoa, o universo muda; cada vez que estende a mão e toca um coração ou uma vida, o mundo se transforma; a cada gentileza e serviço, visto ou não visto, seus propósitos são realizados e nada jamais será igual.

William P. Young em A Cabana

Agradecimentos

Agradeço primeiramente, e acima de tudo, a Deus pela força e por mostrar-me sempre o melhor caminho;

À minha mãe pelo o amor, carinho e dedicação;

À minha pequena, mas incrível família, agradeço pelo apoio e compreensão;

Agradeço ao Professor Roberto da Costa Quinino pelo incentivo e por acreditar no meu potencial;

Aos mestres, agradeço pelos ensinamentos valiosos;

Agradeço, enfim, a todos que, de alguma forma, contribuíram para o presente trabalho e pela formação desse humilde cidadão.

RESUMO

Este trabalho discute o uso de uma função de verossimilhança na estimação da proporção de itens conformes na presença de erros de classificação e de classificações repetidas. Para investigar a qualidade dos estimadores bayesianos foi construído um estudo comparativo entre a técnica proposta e um modelo descrito por Evans *et al.* (1996), considerando, para isso, a influência do tamanho da amostra, do número de classificações repetidas, dos erros de julgamento, da proporção de conformidade e de duas distribuições *a priori* para os erros: Beta(1,2) e Beta(2,10).

As medidas estatísticas utilizadas para a comparação das metodologias foram a moda e o desvio padrão das respectivas distribuições *a posteriori*, além das amplitudes dos intervalos de credibilidade de máxima densidade *a posteriori* e do fator de Bayes. Os resultados encontrados mostraram que o modelo defendido apresentou desvios padrão e amplitudes dos intervalos de credibilidades inferiores àqueles obtidos pela abordagem concorrente, principalmente quando foi considerada a distribuição *a priori* Beta(1,2). Além disso, constatou-se também o fator de Bayes favoreceu o uso do modelo proposto.

Além disso, foi utilizada uma metodologia de otimização multiobjetivo para obter os valores ótimos do tamanho da amostra (n), do número de classificações repetidas (m) e do critério (a) da classificação final dos itens inspecionados após as m classificações. Para realizar essa tarefa foi sugerido um algoritmo genético híbrido multiobjetivo. Exemplos numéricos ilustraram a metodologia proposta e os resultados encontrados podem ser utilizados para ajudar ao processo decisório.

Palavras Chaves: Dados Dicotômicos, Erros de Classificação, Inferência bayesiana, Fator de Bayes, algoritmos genéticos, Otimização Multiobjetivo.

ABSTRACT

This paper discusses the use of a likelihood function for estimating the true proportion of conforming items in the presence of misclassification and repeated classifications. To investigate the quality of Bayes estimators, a comparative study of the technical proposal and a model described by Evans *et al.* (1996) was constructed, varying the sample size, the number of repeated classifications, the inspection errors, the proportion of conformity and two type of prior error distributions: Beta(1,2) and Beta(2,10).

The performances were made by some statistical measurement as: the mode and standard deviation of their posterior distributions, the length of credible intervals of maximum posterior density and the Bayes factor. The results showed that the model proposed presented standard deviations and length of credible intervals smaller than those obtained by the concurrent approach, mainly in the presence of prior distribution Beta(1,2). Furthermore, there is also the Bayes factor favored the use of the proposed model.

In addition, a multi-objective optimization method is used for to obtain the set optimum parameters: sample size (n), the number of repeated classifications (m) and the criterion (a) for the final classification of the inspected item after m classifications. To accomplish this task a hybrid multiobjective genetic algorithm is proposed. Numeric examples illustrate the methodology proposed and the results can be used to help decision-making process.

Key Words: Dichotomous Data, Inspection Errors, Bayesian Inference, Bayes Factor, Genetic Algorithms, Multiobjective Optimization

SUMÁRIO

LISTA DE TABELAS	9
LISTA DE FIGURAS	11
LISTA DE FIGURAS	11
LISTA DE SÍMBOLOS	13
CAPÍTULO 1: INTRODUÇÃO	15
1.1. CONTROLE DE QUALIDADE PARA ATRIBUTOS	15
1.2. JUSTIFICATIVA	17
1.3. MOTIVAÇÃO	18
1.4. OBJETIVOS DO TRABALHO	19
1.4.1. OBJETIVOS GERAIS	19
1.5. ESTRUTURA DA TESE	20
CAPÍTULO 2: REVISÃO DA LITERATURA	21
2.1. REVISÃO DA LITERATURA	21
CAPÍTULO 3: REFERENCIAL TEÓRICO	26
3.1. CLASSIFICAÇÃO DICOTÔMICA COM ERROS	26
3.2. CLASSIFICAÇÕES REPETIDAS	28
CAPÍTULO 4: ESTIMAÇÃO DA PROPORÇÃO DE ITENS CONFORMES NA PRESENÇA DE ERROS E DE CLASSIFICAÇÕES REPETIDAS	35
4.1. INTRODUÇÃO	35
4.2. FUNÇÃO DE VEROSSIMILHANÇA COMPLETA	37
4.3. FUNÇÃO DE VEROSSIMILHANÇA ORDENADA	38
4.3.1. DENSIDADES CONJUNTA E MARGINAL <i>A POSTERIORI</i>	47
4.4. INTERVALO DE CREDIBILIDADE DE MÁXIMA DENSIDADE <i>A POSTERIORI</i>	54
4.5. COMPARAÇÃO DOS MODELOS VIA FATOR DE BAYES	55
4.6. APLICAÇÃO A EXEMPLOS NUMÉRICOS	60
4.7. DELINEAMENTO DO PROCESSO DE SIMULAÇÃO	69
4.8. RESULTADOS NUMÉRICOS	73

CAPÍTULO 5: OTIMIZAÇÃO MULTIOBJETIVO VIA METAHEURÍSTICAS HÍBRIDAS.....	91
5.1. SISTEMAS HÍBRIDOS INTELIGENTES	91
5.2. O PROBLEMA DE OTIMIZAÇÃO MULTIOBJETIVO.....	93
5.2.1. DESCRIÇÃO DO PROBLEMA.....	93
5.2.2. MATRIZ-TRABALHO	100
5.2.3. FUNÇÃO OBJETIVO	101
5.3. ALGORITMO GENÉTICO HÍBRIDO MULTIOBJETIVO PROPOSTO	104
5.3.1. DEFINIÇÃO DO CROMOSSOMO	105
5.3.2. ESTRATÉGIA DO ELITISMO.....	107
5.3.3. CLASSIFICAÇÃO DA POPULAÇÃO.....	109
5.3.4. PRESERVAÇÃO DA DIVERSIDADE DA POPULAÇÃO E CÁLCULO DO <i>FITNESS</i>	112
5.3.5. SELEÇÃO.....	116
5.3.6. MUTAÇÃO	118
5.3.7. RECOMBINAÇÃO	122
5.3.8. BUSCA LOCAL MULTIOBJETIVO.....	124
5.3.9. PSEUDOCÓDIGO DO ALGORITMO GENÉTICO HÍBRIDO MULTIOBJETIVO	132
5.4. APLICAÇÃO A UM EXEMPLO NUMÉRICO	135
5.5. DELINEAMENTO DO PROCESSO DE SIMULAÇÃO	138
5.6. RESULTADOS NUMÉRICOS	143
CAPÍTULO 6: CONCLUSÕES E EXTENSÕES	147
6.1. CONCLUSÕES	147
6.2. EXTENSÕES.....	150
REFERÊNCIAS BIBLIOGRÁFICAS	152

LISTA DE TABELAS

<i>TABELA 3.1: Classificações repetidas de n itens m vezes cada com o critério de $m/2$-----</i>	29
<i>TABELA 3.2: Realização de m classificações repetidas para n itens -----</i>	33
<i>TABELA 4.1: Dados amostrais simulados para o estudo apresentado por Garcia et al. (1997) para m=3, 7 e 11 classificações repetidas, $\epsilon_1=\epsilon_2=0,05$, $p=0,8667$ e distribuições a priori independentes $Beta(1,2)$ para os erros de julgamento-----</i>	61
<i>TABELA 4.2: Estatísticas inferenciais para cada grupo de classificações repetidas -----</i>	62
<i>TABELA 4.3: Frequência Observada de Respostas Positivas (k) entre os Testes Serológicos Repetidos (m) para o Antígeno N da cidade de Hiroshima -----</i>	64
<i>TABELA 4.4: Distribuições a priori para os modelos que serão simulados -----</i>	65
<i>TABELA 4.5: Medidas Estatísticas para a distribuição a posteriori de p -----</i>	68
<i>TABELA 4.6: Valores modais e desvios padrão médios para os modelos proposto e FVC para n=50, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições a priori.-----</i>	75
<i>TABELA 4.7: Valores modais e desvios padrão médios para os modelos proposto e FVC para n=100, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições a priori.-----</i>	76
<i>TABELA 4.8: Valores modais e desvios padrão médios para os modelos proposto e FVC para n=150, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições a priori.-----</i>	77
<i>TABELA 4.9: Tamanhos médios dos intervalos de 95% de credibilidade para os modelos propostos e FVC e Fator de Bayes para n=50, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições a priori. -----</i>	83
<i>TABELA 4.10: Tamanhos médios dos intervalos de 95% de credibilidade para os modelos propostos e FVC e Fator de Bayes para n=100, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições a priori. -----</i>	84
<i>TABELA 4.11: Tamanhos médios dos intervalos de 95% de credibilidade para os modelos propostos e FVC e Fator de Bayes para n=150, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições a priori. -----</i>	85
<i>TABELA 5.1: Total de experimentos possíveis que podem ser gerados a partir da definição dos parâmetros de entrada n, m e a-----</i>	96

<i>Tabela 5.1:</i>	<i>Frequência Observada de Respostas Positivas (k) entre os Testes Serológicos Repetidos (m) para o Antígeno M da cidade de Nagasaki</i> -----	135
<i>Tabela 5.2:</i>	<i>Valores ótimos do tamanho da amostra, o número de classificações repetidas e o critério para a classificação final de cada paciente</i> -----	137
<i>Tabela 5.3:</i>	<i>Valores ótimos de n, m, e a para os parâmetros de entrada da Estratégia 2, considerando uma variância a posteriori máxima de 0,00030 e um tamanho máximo 0,010 para um intervalo MDP de 95% de credibilidade.</i> -----	145

LISTA DE FIGURAS

<i>FIGURA 3.1: Possíveis eventos em uma única classificação ($m=1$) na presença de erros de classificação</i> -----	27
<i>FIGURA 4.1: Funções densidade a posteriori de p para os dados gerados com $n=(100,200)$; $m=5$; $p=0,85$; $\epsilon_1=\epsilon_2=0,05$ e distribuições a priori $U(0,1)$ para os erros de classificação</i> -----	38
<i>FIGURA 4.2: Ilustração do processo de ordenação da amostra pela coluna de classificação final</i> -----	39
<i>FIGURA 4.3: Ilustração de um exemplo fictício de classificação de itens</i> -----	40
<i>FIGURA 4.4: Localização espacial das medidas amostrais calculadas</i> -----	41
<i>FIGURA 4.5: Subdivisão do grupo de classificações finais conformes</i> -----	43
<i>FIGURA 4.6: Obtenção dos limites $M0$ e $M1$</i> -----	45
<i>FIGURA 4.7: Modelagem dos erros de classificação por meio da distribuição beta</i> -----	48
<i>FIGURA 4.8: Ilustração do método de integração de Monte Carlo para o caso unidimensional</i> -----	50
<i>FIGURA 4.9: Obtenção da densidade marginal a posteriori de p via Integração de Monte Carlo</i> -----	52
<i>FIGURA 4.10: Distribuição a posteriori de p para o exemplo numérico apresentado por Garcia et al.</i> -----	63
<i>FIGURA 4.11: Distribuições a priori consideradas no estudo comparativo</i> -----	66
<i>FIGURA 4.12: Distribuições a posteriori de p para nove propostas de modelagem aplicadas ao estudo numérico apresentado por Fujisawa e Izume (2000)</i> -----	67
<i>FIGURA 4.13: Algoritmo de simulação para a comparação dos modelos FVO e FVC</i> -----	71
<i>FIGURA 4.14: Tarefas realizadas para a comparação dos modelos FVO e FVC</i> -----	73
<i>FIGURA 4.15: Comparação dos modelos FVO e FVC por meio do desvio padrão médio, segundo a quantidade de informação a priori, mas controlando pelo tamanho da amostra e pelo número de classificações repetidas</i> -----	78
<i>FIGURA 4.16: Comparação dos modelos FVO e FVC por meio do desvio padrão, segundo a quantidade de informação a priori, mas controlando pelo tamanho da amostra e pela proporção de itens conformes</i> -----	79
<i>FIGURA 4.17: Comparação dos modelos FVO e FVC por meio do desvio padrão, segundo a quantidade de informação a priori, mas controlando pelo tamanho da amostra e pelos erros de classificação</i> -----	80
<i>FIGURA 4.18: Distribuição da razão das amplitudes dos intervalos de 95% de credibilidade do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação a priori, mas controlando pelo tamanho da amostra e pelo número de classificações repetidas</i> -----	86

<i>FIGURA 4.19: Distribuição da razão das amplitudes dos intervalos de 95% de credibilidade do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação a priori, mas controlando pelo tamanho da amostra e pela proporção de itens conformes</i> -----	87
<i>FIGURA 4.20: Distribuição da razão das amplitudes dos intervalos de 95% de credibilidade do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação a priori, mas controlando pelo tamanho da amostra e pelos os erros de classificação</i> -----	88
<i>FIGURA 4.21: Distribuição do fator de Bayes do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação a priori, mas controlado pelo tamanho da amostra e pelo número de classificações repetidas</i> -----	90
<i>FIGURA 5.1: Tarefas realizadas para o cálculo do vetor ótimo (n, m, a)</i> -----	105
<i>FIGURA 5.2: Exemplo fictício da relação de dominância</i> -----	107
<i>FIGURA 5.3: Relação entre os espaços factíveis das variáveis e dos objetivos</i> -----	108
<i>FIGURA 5.4: Exemplo do método de classificação dos conjuntos de soluções de acordo com o valor do fitness</i> ----	110
<i>FIGURA 5.5: Algoritmo Fast Nondominated Sort criado por Deb et al. (2000)</i> -----	111
<i>FIGURA 5.6: Algoritmo de cálculo da distância de aglomeração</i> -----	113
<i>FIGURA 5.7: Cálculo das distâncias de aglomeração pelos métodos padrão e normalizado</i> -----	114
<i>FIGURA 5.8: Algoritmo para o cálculo do fitness no AGHM</i> -----	115
<i>FIGURA 5.9: Cálculo do fitness após obtenção das distâncias de aglomeração normalizadas</i> -----	115
<i>FIGURA 5.10: Estrutura geral de um algoritmo genético tradicional</i> -----	116
<i>FIGURA 5.11: Exemplo do método de agrupamento de Shaffer</i> -----	117
<i>FIGURA 5.12: Função densidade da distribuição normal assimétrica para cinco valores de λ</i> -----	119
<i>FIGURA 5.13: Ilustração gráfica do método de busca local multiobjetivo – BLM</i> -----	125
<i>FIGURA 5.14: Modelo de uma rede neural do tipo PMC</i> -----	128
<i>FIGURA 5.15: Algoritmo para atualizar a população P'_t por meio do algoritmo de busca local multiobjetivo</i> -----	130
<i>FIGURA 5.16: Atualização da população P'_t a partir do conjunto DBL obtido pela busca local</i> -----	131
<i>FIGURA 5.17: Visão geral da implementação do algoritmo genético híbrido multiobjetivo</i> -----	134

LISTA DE SÍMBOLOS

n	Tamanho da amostra
m	Número de classificações repetidas
a	Ponto de corte para a classificação final do item após as m classificações
p	Probabilidade de um item ser conforme
e_1	Probabilidade de classificar um item como não-conforme, sabendo que esse é conforme
e_2	Probabilidade de classificar um item como conforme, sabendo que esse é não-conforme
FVC	Função de verossimilhança completa
FVO	Função de verossimilhança ordenada
C_{ij}	Variável aleatória Bernouli correspondendo a j -ésima classificação do i -ésimo item, sendo 0 item não-conforme e 1 item conforme
F_i	Variável aleatória Bernouli correspondendo à classificação final do i -ésimo item utilizando o ponto de corte a
E_i	Variável aleatória Bernouli correspondendo ao estado real do i -ésimo item
$\lfloor x \rfloor$	É o maior inteiro inferior a x
$B\left(\frac{m}{2}; m; e_k\right)$	É a distribuição binomial acumulada no ponto $m/2$ com probabilidade de sucesso e_k .
S_i	Total de classificações conformes do i -ésimo item avaliado após as m classificações
n_C	Total de itens que tiveram classificações finais conformes utilizando o ponto de corte a
C	Total de classificações repetidas consideradas como conformes nas mn_C classificações realizadas
$n_{\bar{C}}$	Total de itens que tiveram classificações finais não-conformes utilizando o ponto de corte a
\bar{C}	Total de classificações repetidas consideradas como não-conformes nas $mn_{\bar{C}}$ classificações realizadas
IA	Inteligência artificial
SHI	Sistema híbrido inteligente

AG	Algoritmo genético
AGHM	Algoritmo genético híbrido multiobjetivo
P_t	População no tempo t
D_t	Conjunto de soluções dominantes encontradas até o tempo t
N_{elite}	Total de soluções de elite
E_t	É o conjunto das N_{elite} soluções de elite selecionadas a partir do conjunto de soluções dominantes D_t
SP_t	População de soluções-pais no tempo t
P'_t	População de descendentes (soluções-filhos) no tempo t
P_m	Probabilidade de mutação
P_r	Probabilidade de recombinação
BLM	Algoritmo de busca local multiobjetivo
N_{max}	Número máximo de caminhos a serem explorados pelo algoritmo BLM
N_{iter}	número máximo de iterações do algoritmo BLM
DBL	Conjunto de soluções dominantes do algoritmo BLM
$ DBL $	Total de soluções contidas no conjunto DBL
LC	Conjunto de soluções dominantes não selecionadas para o algoritmo BLM
PMC	Rede neural <i>perceptron</i> com multicamadas

Capítulo 1

INTRODUÇÃO

Este capítulo apresenta os elementos iniciais referentes aos aspectos de justificativa, motivação e objetivos do trabalho, bem como sua organização. É importante que essa leitura seja feita com cuidado, pois as informações aqui discutidas servirão de base teórica e metodológica para a compreensão precisa dos demais capítulos.

1.1. CONTROLE DE QUALIDADE PARA ATRIBUTOS

Todos os dias uma enorme quantidade de informações sobre consumidores, produtos, serviços, sistemas de qualidade e monitoramento está sendo coletada nas empresas, nos departamentos, na indústria, entre outras. O volume e a complexidade dos dados tornam a análise uma tarefa praticamente impossível de ser realizada em intervalos de tempos razoáveis.

Na indústria, em especial, a preocupação é quase que exclusivamente voltada para a qualidade dos processos de produção e para a maximização dos lucros. Quando se fala em qualidade de processos produtivos, entende-se que o objetivo é produzir mais e melhor, ou seja, é produzir de forma eficiente, minimizando os recursos, as matérias-primas e maximizando os retornos financeiros. O sucesso de implementação dessa gestão da qualidade dos processos norteará a forma, a quantidade, a época e as características dos itens que serão produzidos.

Além disso, outro aspecto relacionado ao processo produtivo que tem intrigado os gerentes de produção e os tomadores de decisão é o cálculo do tamanho amostral. Determinar, antecipadamente, qual deverá ser o tamanho da amostra associado a um planejamento experimental industrial, por exemplo, de modo que venha garantir as premissas e as especificações do mercado consumidor, é uma tarefa difícil. E o principal obstáculo é o elevado número de fatores que afetam o cálculo do tamanho da amostra, quer sejam pelas influências advindas dos processos produtivos, quer sejam pelas influências geradas pelas características analíticas do modelo estatístico de avaliação proposto.

Indiferente ao problema que se deseje resolver, sendo ele de cunho inferencial ou de otimização do tamanho amostral, o que se objetiva é a construção de processos de produção mais eficientes e pautados na excelência da qualidade total.

No entanto, a busca por essa excelência exige investimento, conhecimento técnico específico e controle constante no sistema de manufatura. Sob esse aspecto, o uso de ferramentas estatísticas de estimação e de previsão torna-se de fundamental importância no entendimento do *mix* de produção.

Assim, este trabalho discute a implementação de um procedimento de estimação de inferência bayesiana da proporção de itens conformes na presença de erros de classificação com a possibilidade da realização de classificações repetidas. Será discutido também um procedimento de otimização multiobjetivo por meio de métodos metaheurísticos híbridos para determinação do tamanho amostral (n), do número de classificações repetidas (m) e do ponto de corte (a), $1 \leq a \leq m$, para a classificação final de um item após terem sido realizadas as m classificações repetidas do mesmo item.

1.2. JUSTIFICATIVA

As metodologias e as técnicas utilizadas neste trabalho trarão uma contribuição para a comunidade acadêmica e científica, principalmente para àquelas áreas de conhecimento relacionadas à saúde e à indústria. E tal contribuição terá dois enfoques: um inferencial e outro de otimização.

Do ponto de vista inferencial, o modelo estatístico proposto neste trabalho se valerá de sua estrutura analítica para estimar, de maneira eficiente, a proporção de dados dicotômicos contendo erros de classificação.

Além disso, um outro ramo da modelagem estatística que tem motivado muitos pesquisadores é a otimização de funções que envolvam variáveis de decisão. Na grande maioria dos casos, os problemas de otimização se limitam ao campo uniobjetivo, envolvendo, geralmente, o cálculo do tamanho amostral, com várias restrições técnicas para que os mesmos tornem-se de resolução fácil e direta. Entretanto, tais soluções podem não ser indicadas para problemas mais complexos. Esta tese irá usar o mesmo problema de estimação da proporção de itens conformes para desenvolver um procedimento de otimização multiobjetivo, utilizando várias técnicas de otimização, tais como: algoritmo genético, redes neurais artificiais e algoritmos de busca local, para encontrar os valores ótimos dos parâmetros do modelo, que são: o tamanho amostral (n), o número de classificações repetidas (m) e o critério de classificação final (a) de cada item avaliado após as m classificações.

Essa metodologia pode ser considerada original no meio científico-industrial, pois são raros os textos que tratam do problema de otimização multiobjetivo via sistemas inteligentes evolutivos e que consideram variáveis de decisão (n, m, a).

1.3. MOTIVAÇÃO

Vários autores em diversos artigos demonstraram a preocupação em determinar a distribuição *a posteriori* da proporção de itens conformes (p) na presença de erros de classificação e/ou classificações repetidas. Na maioria dessas obras, a principal conclusão foi que o uso de distribuições *a priori* informativas contribui, de forma mais eficiente, para a construção de uma distribuição *a posteriori* bem mais “comportada” no sentido de evitar situações multimodais.

No entanto, sabe-se que a utilização de distribuições *a priori* traz consigo uma carga muito grande de subjetividade, por parte do especialista, do grau de conhecimento do fenômeno em estudo, assim como das características dos modelos probabilísticos utilizados. É desejável construir distribuições *a posteriori* unimodais e com a menor variabilidade possível mesmo tendo pouco conhecimento sobre as distribuições de probabilidade dos parâmetros envolvidos.

Os problemas estatísticos existentes nas várias áreas do conhecimento, de um modo geral, referem-se à estimação de parâmetros relacionados às variáveis de decisão. Particularmente nas áreas industrial e médica, há um grande número de obras científicas voltadas para o tema e que se concentram especialmente na utilização das técnicas de inferência bayesiana para resolver problemas analíticos de tomada de decisão (QUININO *et al.*, 2005; LOSCHI; FRANCO, 2006; JOSEPH *et al.*, 1995; JOHNSON *et al.*, 1991). A maior dificuldade observada no uso de tais técnicas é a definição das distribuições *a priori* dos parâmetros dos modelos estatísticos empregados, pois quase sempre é necessário recorrer a uma parametrização dessas distribuições *a priori*, deixando-as totalmente específicas e particularizadas ao problema investigado, ou ainda considerar o uso de distribuições *a priori* empíricas. Explicitando melhor, o ideal é obter distribuições *a posteriori* unimodais e com pouca variabilidade, utilizando o mínimo de conhecimento sobre os parâmetros dos modelos. Essa

idéia, na verdade, é a primeira grande motivação deste trabalho e será discutida sob a ótica de uma função de verossimilhança a ser proposta neste texto no Capítulo 4.

Outro grande desafio da tese está descrito na íntegra no Capítulo 5 e refere-se à aplicação de um algoritmo genético híbrido multiobjetivo para obtenção simultânea de um vetor ótimo de parâmetros: tamanho amostral, número de classificações repetidas e ponto de corte utilizado para a classificação final de cada item avaliado na amostra. Esse procedimento tem um papel importante na área de otimização de inferência bayesiana, pois abordará o mesmo problema comentado anteriormente de uma maneira simples, intuitiva e com um enfoque industrial. Além daquele já citado, o maior sentimento motivador dessa tese se concentra na capacidade de generalização e de adaptação do programa proposto aos problemas existentes em outros campos de conhecimento.

1.4. OBJETIVOS DO TRABALHO

1.4.1. OBJETIVOS GERAIS

Esta produção científica será fundamentada em dois grandes pilares teóricos, a citar:

1. Construir uma função de verossimilhança a partir da proposição de um modelo em que é implementado um procedimento de inferência bayesiana para a estimação da proporção de itens conformes na presença de erros de classificação e de classificações repetidas. Esse modelo será comparado com a metodologia desenvolvida por Evans *et al.* (1996);
2. Utilizar a função de verossimilhança proposta no item 1, para desenvolver um algoritmo genético híbrido multiobjetivo para estimar concomitantemente os valores ótimos do tamanho amostral, do número de classificações repetidas e do ponto de corte usado para a classificação final dos itens após serem realizadas todas as classificações repetidas.

1.5. ESTRUTURA DA TESE

Esta tese está organizada em seis capítulos. O capítulo inicial, conforme já visto nas seções anteriores, trata das informações preliminares e de preparação do entendimento global dos objetivos do trabalho. O segundo capítulo se concentrará numa revisão bibliográfica das principais obras e autores no que tange aos preceitos teóricos da tese. Já o Capítulo 3 será dedicado ao desenvolvimento de todo o referencial teórico do trabalho, desde o surgimento das primeiras ideias até as propostas mais atuais.

O Capítulo 4, por sua vez, abordará o primeiro objetivo geral desta tese, apresentando e discutindo a função de verossimilhança proposta. Os resultados de simulação foram realizados para a validação do modelo. Enquanto que o Capítulo 5 tratará o tema *otimização multiobjetivo* do vetor (n, m, a) , perpassando por algumas técnicas evolutivas, tais como redes neurais artificiais e algoritmos genéticos. Por fim, as considerações finais e as sugestões de estudos futuros nortearão o Capítulo 6, e servirão, sobretudo, para vislumbrar as possibilidades de ampliação do estudo realizado.

Capítulo 2

REVISÃO DA LITERATURA

A revisão da literatura e a evolução histórica das metodologias teóricas que serão utilizadas nesta tese são os objetivos deste capítulo.

2.1. REVISÃO DA LITERATURA

Um sistema de controle, visto de forma ampla, pode ser considerado como um programa planejado e documentado compreendendo pessoal, máquinas, equipamentos, instrumentos, procedimentos e padrões, tudo com o objetivo de proporcionar uma melhoria contínua da qualidade. A inspeção direta de um produto ou serviço é uma atividade necessária dentro do sistema de controle e, para tanto, é fundamental que o sistema de inspeção a ser empregado seja eficiente. Dessa forma, evita-se que sejam tomadas decisões equivocadas, tais como: utilizar itens não-conformes ou descartar itens conformes.

Quando a inspeção baseia-se em atributos, isto é, uma inspeção na qual uma ou mais características da qualidade são analisadas, a unidade de produto ou serviço correspondente, que parte de um lote, batelada ou população, pode ser classificada simplesmente como conforme ou não-conforme em relação a uma dada característica. Baseando-se nas frequências de conformidade observadas, determina-se a necessidade ou não de interferência no processo produtivo.

Um dos grandes problemas em estabelecer um sistema de inspeção para produtos classificados como conforme ou não-conforme é a falta de informações sobre a eficiência do sistema de classificação utilizado. Dois tipos de erros podem ocorrer: o primeiro, conhecido como erro tipo I (e_1), é a probabilidade de classificar um item originalmente conforme como não-conforme; e o segundo, chamado de erro tipo II (e_2), é a probabilidade classificar um item como não-conforme, sabendo que esse é originalmente conforme.

Os efeitos de se ignorar os erros de classificação foram pioneiramente estudados por Bross (1954), que mostrou que estimações clássicas baseadas em proporções são viciadas. Em termos práticos, a estimativa da proporção de interesse poderia estar comprometida e decisões poderiam ser tomadas equivocadamente. Ele também investigou a perda de poder em tabelas de contingência 2×2 . Outros autores estenderam o trabalho de Bross considerando os efeitos dos erros de classificação em tabelas de contingência generalizadas e na derivação de planejamentos ótimos. Exemplos desses estudos são os trabalhos desenvolvidos por Fleiss (1973) e Kleinbaum, Kupper e Morgenstern (1982) que referenciaram as fontes de pesquisas de campo como um dos principais problemas das classificações binárias.

Além do efeito dos erros de classificação em análises de dados categóricos no que se refere à estimação da proporção de sucesso (itens conformes, por exemplo), Bross (1954) também mostrou que o conhecimento das taxas de erros de classificação é uma medida necessária para a correção do vício resultante da estimação baseada nas proporções observadas na amostra. No início da década de 70 do século XX, Tenenbein (1970, 1971, 1972) propôs obter essa informação por meio de um processo de amostragem dupla, no qual uma parte da amostra foi utilizada para estimar a proporção de sucesso por meio de classificador de atributos e a outra parte foi utilizada para realizar validação cruzada desse classificador. Mas o que fazer quando não se dispõe de um classificador infalível de itens ou quando o seu uso torna o processo de controle de qualidade muito dispendioso, que é o caso

dos problemas industriais? Uma alternativa é a abordagem bayesiana em que se assume o conhecimento da distribuição *a priori* dos parâmetros de classificação incorreta e incorpora à informação amostral via regra de Bayes.

Lindsey (1985) apresentou uma forma matemática de um modelo de misturas de binomiais para estimar a proporção de itens defeituosos considerando a presença de erros de classificação. Já Gaba e Winkler (1992) discutiram modelos bayesianos aplicados a experimentos com uma única classificação de cada item na presença de erros de classificação.

Um pouco antes da publicação da obra de Gaba e Winkler (1992), Johnson e Kotz (1991) e Johnson *et al.* (1991) discutiram a questão de avaliação de atributos em controle de qualidade industrial. Além disso, outros autores, como Greenberg e Stokes (1995), Evans *et al.* (1996) e Gustafson (2003) enfatizaram que os erros de classificação não podem ser negligenciados, pois eles podem comprometer seriamente o processo de inspeção e tomada de decisão. Evans *et al.* (1996) apresentaram um estudo de inferência da proporção sujeita a erros de classificação e avaliações repetidas. Esse estudo será a base de comparação dessa tese. Além disso, Burke *et al.* (1995), Gramopadhye *et al.* (1995), Kemp e Kemp (1988) também relatam exemplos de erros de diagnóstico que influenciam criticamente o processo de avaliação. Neste sentido, o controle de qualidade por atributos depende da eficiência do sistema de inspeção adotado.

Em geral, as metodologias existentes requerem o conhecimento do estado real dos itens avaliados para viabilizar a análise dos erros tipo I e tipo II. Barry e Christopher (1981), Johnson e Funke (1980) e Sylla (2002) apresentaram métodos para julgar o sistema de inspeção quando a classificação real dos itens é conhecida. Quando não há tal conhecimento, Johnson *et al.* (1991) mostraram que não é possível estimar os erros de classificação sem a utilização de informações adicionais. Por sua vez, Joseph *et al.* (1995) propuseram métodos bayesianos para incorporar os

erros de classificação. As funções de densidade marginal *a posteriori* dos erros foram obtidas, tornando possível julgar o sistema de inspeção. No entanto, tais métodos dependem fortemente da especificação de uma distribuição *a priori* para os parâmetros, que necessita ser muito informativa para obtenção de uma distribuição *a posteriori* com poucos valores modais.

Obter distribuições *a priori* muito informativas pode não ser possível, pois a necessidade de julgar o sistema de inspeção tem como principal motivação a alta incerteza sobre a magnitude dos erros de classificação cometidos, contradizendo a especificação de uma distribuição *a priori* muito informativa.

Kutner *et al.* (1996) utilizaram modelos de efeitos aleatórios para avaliar a precisão de testes diagnósticos. Nessa mesma linha de pesquisa, Fujisawa e Izume (2000) empregaram modelos de classes latentes com respostas binárias para investigar a confiabilidade de testes de diagnóstico de falso positivo e falso negativo, ou seja, para avaliar a magnitude dos erros. Já Quinino *et al.* (2005), em uma abordagem bayesiana para a avaliação de inspetores em classificações dicotômicas, utilizaram distribuições *a priori* dispersas para os parâmetros do modelo. Classificaram cada item avaliado m vezes e consideraram o estado real do item avaliado como conforme se a maioria dessas classificações fosse conforme. Discutiram um modelo para cada erro de classificação separando as peças consideradas conformes e não conformes. Com a reconstrução do estado real dos itens, procedeu-se uma análise bayesiana usual para cada conjunto de dados. E mais recentemente, Pires (2006) e Moura (2007) apresentaram algumas soluções para estudar a distribuição *a posteriori* da proporção de itens conformes (p) na presença de erros de diagnóstico. Apesar de todos os esforços mencionados para se estudar o efeito dos erros de classificação em processos decisórios dicotômicos, é eminente a necessidade de se construir modelos investigativos mais eficientes e que utilizem a menor quantidade possível de restrições ou de metodologias empíricas, sendo essa a maior motivação para esse tema.

Além do estudo da proporção de conformidade de classificadores para atributos, há um interesse em se investigar o uso de métodos de otimização combinatória multiobjetivo para a determinação dos valores ótimos de variáveis de decisão, tais como: o tamanho da amostra (n), o número de classificações repetidas (m) e o critério (a) para a classificação final do item avaliado. Nesse sentido, Rahme *et al.* (2000) propuseram um método incremental para a obtenção do tamanho amostral ótimo. Essa obra, embora pioneira, representou a resolução de um problema de otimização uniobjetivo. O que se deseja, no entanto, é a construção de metodologias capazes de resolverem problemas multiobjetivos.

Embora quase não existiam obras direcionadas ao tratamento específico de experimentos de Bernoulli, Coello (2000), Ehrgott (2000) e Jones *et al.* (2002) discutiram algumas técnicas para estudar problemas de otimização multiobjetivo, principalmente utilizando modelos de computação evolutiva e genética. Os algoritmos genéticos e as redes neurais são os principais representantes dessas estruturas computacionais. Mais especificamente, os algoritmos genéticos, idealizados por John Holland (1975), têm sido muito empregados para resolver os mais diversos problemas nas áreas de engenharia, médica, de segmentação de mercado, de robótica, de previsão do tempo e industrial. Conhecido por imitar a teoria de evolução das espécies, os algoritmos genéticos configuram uma ferramenta versátil e adaptável a qualquer problema de estudo, mas não infalível. Por isso, é comum em muitas situações práticas a construção de sistemas híbridos inteligentes que unem duas ou mais técnicas de otimização. Essa estratégia é utilizada, especialmente, para suprir as deficiências das técnicas individuais.

Assim, é também meta dessa tese a obtenção do vetor ótimo (n, m, a) a partir da distribuição *a posteriori* da proporção de itens conformes por meio de um algoritmo genético híbrido multiobjetivo.

Capítulo 3

REFERENCIAL TEÓRICO

Este capítulo possui como objetivo desenvolver o material teórico e metodológico que fomentará todos os capítulos subseqüentes.

3.1. CLASSIFICAÇÃO DICOTÔMICA COM ERROS

Considere que uma amostra aleatória de n itens seja retirada de um processo de produção e que cada item seja julgado como conforme ou não-conforme, sendo p a probabilidade de conformidade. O número de itens conformes observados (X) possui distribuição binomial com parâmetros (n, p) , ou simbolicamente, $X \sim Bin(n, p)$. Contudo, a existência de erros de classificação no sistema implica numa modificação dessa função de probabilidade.

Seja e_1 a probabilidade de que um item conforme seja erroneamente classificado como não-conforme, e seja e_2 a probabilidade de que um item não-conforme seja classificado como conforme. Essas duas probabilidades são conhecidas como erros Tipo I e II, respectivamente. Então, a probabilidade de que um item seja classificado como conforme é dada por $q = p(1 - e_1) + (1 - p)e_2$, conforme ilustra a FIGURA 3.1.

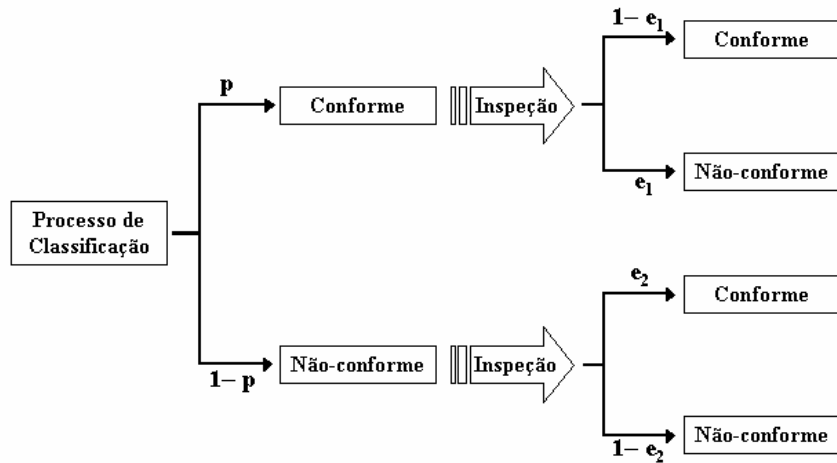


FIGURA 3.1: Possíveis eventos em uma única classificação ($m=1$) na presença de erros de classificação

Desse modo, na presença de erros de classificação, a variável aleatória X tem distribuição binomial com parâmetro q ao invés de p . A dificuldade encontrada em tais análises pode ser melhor identificada quando se tenta obter o estimador de máxima verossimilhança. A função de verossimilhança para esse caso, que apresenta erros de classificação, pode ser escrita como $L(x | n, q) \propto q^x (1 - q)^{n-x}$. Esta função é maximizada para todos os pontos (p, e_1, e_2) , tais que $p(1 - e_1) + (1 - p)e_2 = \frac{x}{n}$, mas pode apresentar mais de uma solução (GABA; WINKLER, 1992).

Assim, no sentido de resolver esse problema, vários métodos clássicos têm sido sugeridos, como apresentado em Evans *et al.* (1996) e uma revisão pode ser encontrada em Johnson *et al.* (1991). Em geral, os métodos propostos baseiam-se em planos amostrais alternativos para a estimação preliminar dos erros de classificação. Do ponto de vista bayesiano, Gaba e Winkler (1992) consideraram um procedimento que requer o uso de uma distribuição *a priori* informativa. Isso pode ser considerado como uma restrição porque, em muitos casos, essa informação não está disponível. Eles também verificaram que o uso de distribuições *a priori* independentes não-informativas e

uniformes entre 0 e 1, para cada parâmetro p , e_1 e e_2 , produzem uma média da distribuição *a posteriori* igual a $1/2$, independente do plano amostral, bem como para todo ponto $(p, e_1$ e $e_2)$, tal que $p(1 - e_1) + (1 - p)e_2 = \frac{x}{n}$.

3.2. CLASSIFICAÇÕES REPETIDAS

Suponha agora, na estrutura inicial de classificação (FIGURA 3.1), que cada item da amostra aleatória de tamanho n seja classificado m vezes, independentemente, como conforme ou não-conforme. Nesta tese o número de classificações repetidas (m) será considerado fixo para todos os itens. Seja C_{ij} ($i=1, 2, \dots, n; j=1, 2, \dots, m$) uma variável aleatória Bernoulli(q) correspondente à j -ésima classificação do i -ésimo item, em que $C_{ij} = 0$ representa um item não-conforme. Dessa forma, $C_{34} = 1$ significa que o terceiro item da amostra foi classificado como conforme em sua quarta classificação.

Seja F_i uma variável aleatória Bernoulli que denota a classificação final do i -ésimo item após terem sido realizadas as m classificações. Claramente, a classificação final dependerá do critério que se deseja assumir, ou seja, da escolha do ponto de corte. Se for considerando o critério da maioria, ou seja, se for utilizado um ponto de corte $m/2$, então, $F_i = 1$ se e somente se

$$\sum_{j=1}^m C_{ij} > \frac{m}{2} \text{ para } i=1, 2, \dots, n. \text{ A escolha de um número ímpar de classificações repetidas elimina as}$$

possibilidades de empates e, conseqüentemente, facilita o método de classificação final de cada item.

Uma forma alternativa de classificar o item seria avaliá-lo até que ocorrer $\frac{m}{2} + 1$ classificações

conformes. Esse critério visa, principalmente, a redução dos custos de produção. A TABELA 3.1 ilustra esse procedimento de classificação.

TABELA 3.1: Classificações repetidas de n itens m vezes cada com o critério de $m/2$

Item	Classificações Repetidas (C_{ij})					Classificação Final (F_i)
	1	2	3	m	
1	C_{11}	C_{12}	C_{13}	C_{1m}	F_1
2	C_{21}	C_{22}	C_{23}	C_{2m}	F_2
3	C_{31}	C_{32}	C_{33}	C_{3m}	F_3
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	C_{n1}	C_{n2}	C_{n3}	C_{nm}	F_n

As variáveis $F_i, i=1, 2, \dots, n$, são independentes e identicamente distribuídas. Dessa maneira, definindo outra variável aleatória Bernoulli E_i como sendo o estado real do i -ésimo item, em que $E_i = 1$ representa um item originalmente conforme, existe um interesse de estimar $P(E_i = 1) = p$, de modo que $e_1 = P(C_{ij} = 0 | E_i = 1)$ e $e_2 = P(C_{ij} = 1 | E_i = 0)$ denotam as probabilidades de ocorrência dos erros tipo I e II, respectivamente. Utilizando, além disso, o teorema da probabilidade total, mostra-se que $P(F_i = 1)$ pode ser escrita por

$$\begin{aligned}
 P(F_i = 1) &= P(F_i = 1, E_i = 1) + P(F_i = 1, E_i = 0) = \\
 &= P(E_i = 1)P(F_i = 1 | E_i = 1) + P(E_i = 0)P(F_i = 1 | E_i = 0)
 \end{aligned}
 \tag{3.1}$$

Observe na equação 3.1 que $P(F_i = 1 | E_i = 1)$ é a probabilidade do i -ésimo item ser classificado como conforme após as m classificações repetidas, sabendo que o seu estado real é conforme, refletindo, assim, a influência de e_1 no cálculo dessa quantidade. Em outras palavras, considerando o critério da maioria ($m/2$) adotado, essa probabilidade é calculada como segue:

$$P(F_i = 1 | E_i = 1) = 1 - \sum_{x=0}^{\lfloor m/2 \rfloor} \binom{m}{x} (1 - e_1)^x (e_1)^{m-x} \quad (3.2)$$

em que $\lfloor m/2 \rfloor$ denota o maior inteiro menor ou igual a $m/2$.

Analogamente, define-se $P(F_i = 1 | E_i = 0)$ como sendo a probabilidade do i -ésimo item ser classificado como conforme após as m classificações repetidas, sabendo que o seu estado real é não-conforme. Nesse caso, é o erro tipo II que contribui para a probabilidade de uma classificação equivocada acontecer e que pode ser calculada como

$$P(F_i = 1 | E_i = 0) = 1 - \sum_{x=0}^{\lfloor m/2 \rfloor} \binom{m}{x} (e_2)^x (1 - e_2)^{m-x} \quad (3.3)$$

Uma vez que, por definição, $P(E_i = 1) = p$, é direto reescrever a expressão 3.1 a partir das equações 3.2 e 3.3 como

$$P(F_i = 1) = p \left[1 - \sum_{x=0}^{\lfloor m/2 \rfloor} \binom{m}{x} (1 - e_1)^x (e_1)^{m-x} \right] + (1 - p) \left[1 - \sum_{x=0}^{\lfloor m/2 \rfloor} \binom{m}{x} (e_2)^x (1 - e_2)^{m-x} \right] \quad (3.4)$$

Utilizando a notação de $B\left(\frac{m}{2}; m; e_k\right) = \sum_{x=0}^{\lfloor m/2 \rfloor} \binom{m}{x} (e_k)^x (1 - e_k)^{m-x}$ como a distribuição binomial acumulada até o ponto $m/2$, tal que $\bar{B}\left(\frac{m}{2}; m; e_k\right) = 1 - B\left(\frac{m}{2}; m; e_k\right)$, então a expressão 3.4 pode ser resumida em

$$P(F_i = 1) = p \bar{B}\left(\frac{m}{2}; m; 1 - e_1\right) + (1 - p) \bar{B}\left(\frac{m}{2}; m; e_2\right) \quad (3.5)$$

Para sistemas reais de classificação é razoável se pensar que os erros de classificação e_1 e e_2 influenciem o processo, mas tenham valores baixos, pois não faz sentido construir sistemas classificadores de atributos, cujos erros de julgamento sejam altos. Por isso, Johnson *et al.* (1991) afirmam que é razoável assumir os erros e_1 e e_2 como sendo inferiores a 0,5. Adotando como ponto de corte $a = m/2$, observa-se que o crescimento de m implicará no decréscimo da probabilidade de julgamentos equivocados sobre a conformidade ou não dos produtos. Isto é, à medida que m aumenta, a proporção de conformidade converge para a verdadeira proporção, corroborando, assim, com a indicação do uso de classificações repetidas. Tudo isto pode ser observado de uma forma um pouco mais geral na Proposição 1.

PROPOSIÇÃO 1: Com o aumento de m , a proporção amostral de itens classificados como conformes converge para o valor verdadeiro da proporção p de itens conformes quando $e_1 < 0,5$ e $e_2 < 0,5$.

Prova: Matematicamente, a expressão 3.4 converge para p , quando $m \rightarrow \infty$ e $e_1 < 0,5$ e $e_2 < 0,5$. A demonstração pode ser realizada utilizando a aproximação da distribuição binomial pela distribuição normal, de modo que

$$\begin{aligned} \lim_{m \rightarrow \infty} P(F_i = 1) &= \lim_{m \rightarrow \infty} \left\{ p \left[1 - \Phi \left(\frac{\sqrt{m}(2e_1 - 1)}{2\sqrt{e_1(1 - e_1)}} \right) \right] + (1 - p) \left[1 - \Phi \left(\frac{\sqrt{m}(1 - 2e_2)}{2\sqrt{e_2(1 - e_2)}} \right) \right] \right\} \\ &= p[1 - \Phi(-\infty)] + (1 - p)[1 - \Phi(\infty)] = p \end{aligned} \quad (3.6)$$

Supondo, agora, que em uma amostra aleatória de n itens, r desses foram considerados conformes após terem sido classificados m vezes, então a função de verossimilhança pode ser expressa por

$$L(r | n, m, p, e_1, e_2) \propto \left\{ pB\left(\frac{m}{2}; m; e_1\right) + (1-p)\bar{B}\left(\frac{m}{2}; m; e_2\right) \right\}^r \times \left\{ 1 - pB\left(\frac{m}{2}; m; e_1\right) - (1-p)\bar{B}\left(\frac{m}{2}; m; e_2\right) \right\}^{n-r} \quad (3.7)$$

Note que se $m = 1$, então a equação 3.7 torna-se

$$L(r | n, m, p, e_1, e_2) = [p(1-e_1) + (1-p)e_2]^r [pe_1 + (1-p)(1-e_2)]^{n-r} \quad (3.8)$$

A expressão 3.8 é exatamente a mesma descrita em Gaba e Winkler (1992) e Viana *et al.* (1993), indicando que a expressão 3.7 é uma generalização desses modelos obtida por meio da inclusão de classificações repetidas. Essa generalização foi justamente a principal contribuição de Pires (2006) às ideias iniciadas por Gaba e Winkler (1992). A partir do que foi discutido, Pires (2006) construiu um estudo de inferência bayesiana para a proporção de itens conformes na presença de erros de classificação e de classificações repetidas com uso de distribuições *a priori* empíricas. Em termos práticos, Pires (2006) conseguiu resultados coerentes, mas as distribuições *a posteriori* dos parâmetros estudados continuaram multimodais e com grande variabilidade.

Outra abordagem possível é utilizar o modelo de Lindsey (1985) e Evans *et al.* (1996) para avaliar as distribuições *a posteriori* de p , e_1 e e_2 , abordando a mesma estrutura de dados, como é mostrada a seguir.

Considere, agora, uma variável aleatória S_i como sendo o total de classificações conformes do i -ésimo item avaliado após as m classificações, ou seja, $S_i = \sum_{j=1}^m C_{ij}$. Desse modo, considere uma modificação feita na TABELA 3.1, no sentido de considerar essa nova formulação, tendo a seguinte configuração

TABELA 3.2: Realização de m classificações repetidas para n itens

Item	Classificações Repetidas (C_{ij})					Classificações Conformes (S_i)
	1	2	3	...	m	
1	C_{11}	C_{12}	C_{13}	...	C_{1m}	S_1
2	C_{21}	C_{22}	C_{23}	...	C_{2m}	S_2
3	C_{31}	C_{32}	C_{33}	...	C_{3m}	S_3
\vdots	\vdots	\vdots	\vdots	...	\vdots	\vdots
n	C_{n1}	C_{n2}	C_{n3}	...	C_{nm}	S_n

Assim, a probabilidade de observar-se o vetor $\mathbf{S}=(S_1, S_2, \dots, S_n)$ é dada por

$$P(S = s) = \prod_{i=1}^n \left[p \binom{m}{s_i} (1-e_1)^{s_i} (e_1)^{m-s_i} + (1-p) \binom{m}{s_i} (e_2)^{s_i} (1-e_2)^{m-s_i} \right] \quad (3.9)$$

Essa probabilidade conjunta pode ser entendida como uma função de verossimilhança para os dados do experimento. Sejam $n_k = \sum_{i=1}^n I_{\{s_i=k\}}$, $k=0, 1, \dots, m$ as frequências de cada um dos resultados possíveis das variáveis s_i . Assim, a expressão 3.9 pode ser reescrita como

$$P(S = s) = \prod_{k=0}^m \left[p \binom{m}{k} (1-e_1)^k (e_1)^{m-k} + (1-p) \binom{m}{k} (e_2)^k (1-e_2)^{m-k} \right]^{n_k} \quad (3.10)$$

Desenvolvendo a expressão (3.10) a partir do binômio de Newton e definindo $\boldsymbol{\eta}=(n_1, n_2, \dots, n_m)$, é possível expressar a função de verossimilhança por

$$L(p, e_1, e_2 | m, \boldsymbol{\eta}) \propto \prod_{k=0}^m \sum_{u=0}^{n_k} \binom{m}{u} \left[p(1-e_1)^k (e_1)^{m-k} \right]^u \left[(1-p)(e_2)^k (1-e_2)^{m-k} \right]^{n_k-u} \quad (3.11)$$

A equação 3.11 foi desenvolvida por Moura (2007), que empregou uma análise bayesiana para estimar os erros de classificação e aproximar as distribuições *a posteriori* da proporção de conformidade e dos respectivos erros. Os resultados encontrados foram melhores que àqueles obtidos por Pires (2006), pelo menos, em termos de variabilidade das estimativas dos parâmetros, embora com os mesmos problemas de distribuições *a posteriori* multimodais. Nas obras pesquisadas, no que se refere ao tamanho amostral com abordagem bayesiana para dados dicotômicos na presença de erros de classificação, Dendukuri *et al.* (2004) e Rahme *et al.* (2000) notaram, também, a necessidade imprescindível em especificar uma distribuição *a priori* informativa.

Além da investigação sobre a proporção de itens conformes e erros de classificação, outro interesse está relacionado com os métodos de otimização multiobjetivo que envolvem, por exemplo, uma função objetivo do tipo $f = f(x, y)$, em que x representa a variância da distribuição *a posteriori* de p e y o comprimento de seu intervalo de credibilidade. Assim, a meta seria encontrar o vetor ótimo dos parâmetros de entrada (n, m, a) que, ao mesmo tempo, minimizasse as componentes x e y .

Nessa linha de raciocínio, vários autores, como Coello (2000), Ehrgott (2000) e Jones *et al.* (2002), propuseram técnicas para estudar os problemas de otimização multiojetivo, principalmente utilizando modelos de computação evolutiva, mas quase nenhuma obra está relacionada a dados com características dicotômicas.

Mesmo assim, sabe-se que ainda existe uma grande escassez de material teórico sobre o tema. Por isso, resolveu-se sugerir uma meta-heurística híbrida que envolveu redes neurais, algoritmos genéticos e um método de busca local para estudar problemas de otimização combinatória multiobjetivo de controle de atributos.

Capítulo 4

ESTIMAÇÃO DA PROPORÇÃO DE ITENS CONFORMES NA PRESENÇA DE ERROS E DE CLASSIFICAÇÕES REPETIDAS

Este capítulo é dedicado à definição, discussão e apresentação dos principais resultados referentes à metodologia proposta nesta tese. Uma avaliação de desempenho dessa nova proposta será realizada por meio de comparação com o modelo desenvolvido por Evans *et al.* (1996).

4.1. INTRODUÇÃO

Em várias aplicações e, particularmente, no contexto produtivo, os dados binários estão sujeitos a erros de classificação. Por exemplo, um determinado item de um sistema produtivo, originalmente perfeito, pode ser classificado como defeituoso por algum sistema automático de qualidade. Da mesma forma um item originalmente defeituoso pode ser classificado como perfeito. Bross (1954) foi o primeiro a investigar a influência dos erros de classificação no vício dos estimadores clássicos baseados na proporção. Recentemente outros autores, tais como Evans *et al.* (1996) e Pires (2006), apresentaram soluções para a distribuição *a posteriori* da proporção de conformidade.

Em vista dos conceitos e das formulações que foram apresentados no Capítulo 3, o desenvolvimento deste trabalho se deu com base em uma proposta de uma função de verossimilhança para resolver os problemas discutidos por Lindsey (1985), Gaba e Winkler (1992), Evans *et al.* (1996) e Quinino *et al.* (2005).

Mais especificamente, a meta deste capítulo consiste em estimar, sob a abordagem bayesiana, a proporção de itens conformes de dados dicotômicos na presença de erros de avaliação de atributos e de classificações repetidas, a partir de uma função de verossimilhança proposta, chamada aqui de função de verossimilhança ordenada (FVO).

Portanto, este capítulo irá investigar as características inferenciais da distribuição *a posteriori* da proporção de itens conformes em estudos dicotômicos com erros e classificações repetidas. Esse estudo será desenvolvido a partir da comparação de dois modelos: **(a)** função de verossimilhança Completa (FVC), idealizada por Lindsey (1985) e complementada por Evans *et al.* (1996) e **(b)** função de verossimilhança Ordenada (FVO), proposta defendida nesta tese. O desvio padrão, a amplitude do intervalo de credibilidade e o fator de Bayes de ambos os modelos serão os parâmetros balizadores dessa comparação.

Assim, a seção 4.2 apresentará a função de verossimilhança completa, a seção 4.3 discutirá a função de verossimilhança proposta. Nas seções 4.4 e 4.5 serão abordados a obtenção do intervalo de credibilidade e o cálculo do fator de Bayes, respectivamente. Já na seção 4.6 é exposta uma aplicação das metodologias a dois exemplos numéricos, enquanto que na seção 4.7 é definido o processo de simulação. Por fim, na seção 4.8 será apresentada uma discussão dos resultados obtidos.

4.2. FUNÇÃO DE VEROSSIMILHANÇA COMPLETA

Para a estrutura de dados apresentada no Capítulo 3, foi definida uma função de verossimilhança (equação 3.9) baseada na variável aleatória $S_i = \sum_{j=1}^m C_{ij}$ que denotava o total de classificações conformes. Pioneiramente construída por Lindsey (1985) e reeditada por Evans *et al.* (1996), esse processo de construção tem a característica de utilizar toda a informação amostral, em termos do número de classificações finais conformes de todos os itens avaliados, sendo, por isso, também chamado de modelo de verossimilhança completa.

$$P(S = s) = \prod_{i=1}^n \left[p \binom{m}{s_i} (1 - e_1)^{s_i} (e_1)^{m-s_i} + (1 - p) \binom{m}{s_i} (e_2)^{s_i} (1 - e_2)^{m-s_i} \right] \quad (4.1)$$

A expressão 4.1 também pode ser vista como uma função de verossimilhança para o experimento amostral. Adotando-se distribuições *a priori* $U(0,1)$ independentes para os parâmetros (p, e_1, e_2) a função densidade *a posteriori* de p será proporcional à expressão (4.1). Entretanto, tal procedimento pode resultar em uma distribuição *a posteriori* bimodal e com variância elevada.

Os resultados obtidos por simulação, considerando cinco classificações repetidas, $e_1=0,05$, $e_2=0,05$ e tamanhos amostrais iguais a 100 e 200, gerados para se determinar a distribuição *a posteriori* de p , revelaram os problemas relacionados ao uso da função de verossimilhança adotada (FIGURA 4.1), ou seja, distribuições dispersas e com duas modas. Na verdade não existe prova analítica que justifique a forma dos gráficos, mas partindo da equação (4.1) que determina um modelo de mistura de binomiais, parece que existe uma tendência de que as modas da distribuição *a posteriori* ocorram em torno dos pontos $1-e_1$ e e_2 .

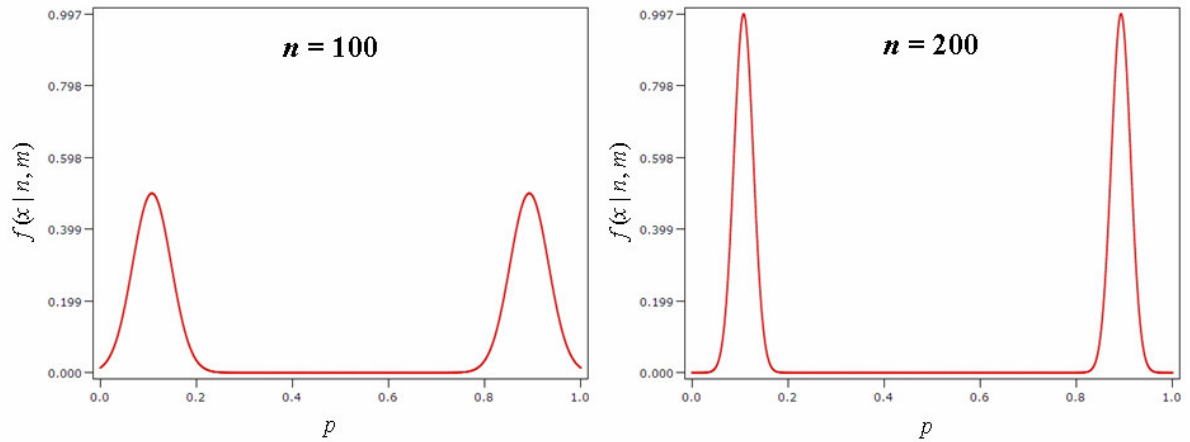


FIGURA 4.1: Funções densidade *a posteriori* de p para os dados simulados com $n=(100,200)$; $m=5$; $p=0,85$; $e_1=e_2=0,05$ e distribuições *a priori* $U(0,1)$ para os erros de classificação

Essas características motivaram a construção de uma nova formulação, chamada de função de verossimilhança ordenada, discutida na próxima seção e que será um dos objetivos desta tese. Por entender que um processo de ordenação aumenta o grau de informação sobre um determinado parâmetro, a expectativa inicial é obter distribuições *a posteriori* com variância inferior àquela conseguida por Evans *et al.* (1996).

4.3. FUNÇÃO DE VEROSSIMILHANÇA ORDENADA

A metodologia aqui proposta se vale da extração de informações adicionais provenientes do mesmo experimento descrito nas seções anteriores. Assim, considere um processo de classificação de n itens em conforme e não-conforme. Seja p a probabilidade de um determinado item ser conforme; em que e_1 é a probabilidade de que um determinado item conforme seja classificado como não-conforme em uma única classificação, e e_2 represente a probabilidade de que um determinado

item não-conforme seja classificado como conforme em uma única classificação. Considere, ainda, como definidas no Capítulo 3, as variáveis aleatórias de Bernoulli C_{ij} e F_i , indicando a j -ésima classificação do i -ésimo item e a classificação final do i -ésimo item utilizando o critério da maioria.

Após a realização do processo de classificação, e conseqüentemente ter obtido o vetor $\mathbf{F} = (F_1, F_2, \dots, F_n)$, toda a matriz de dados é ordenada por esse vetor de classificações finais, o que gerará um novo identificador, dentro de uma nova ordem, para cada item avaliado, como é ilustrado na FIGURA 4.2. Para melhor entendimento do processo, um exemplo é mostrado na FIGURA 4.3.

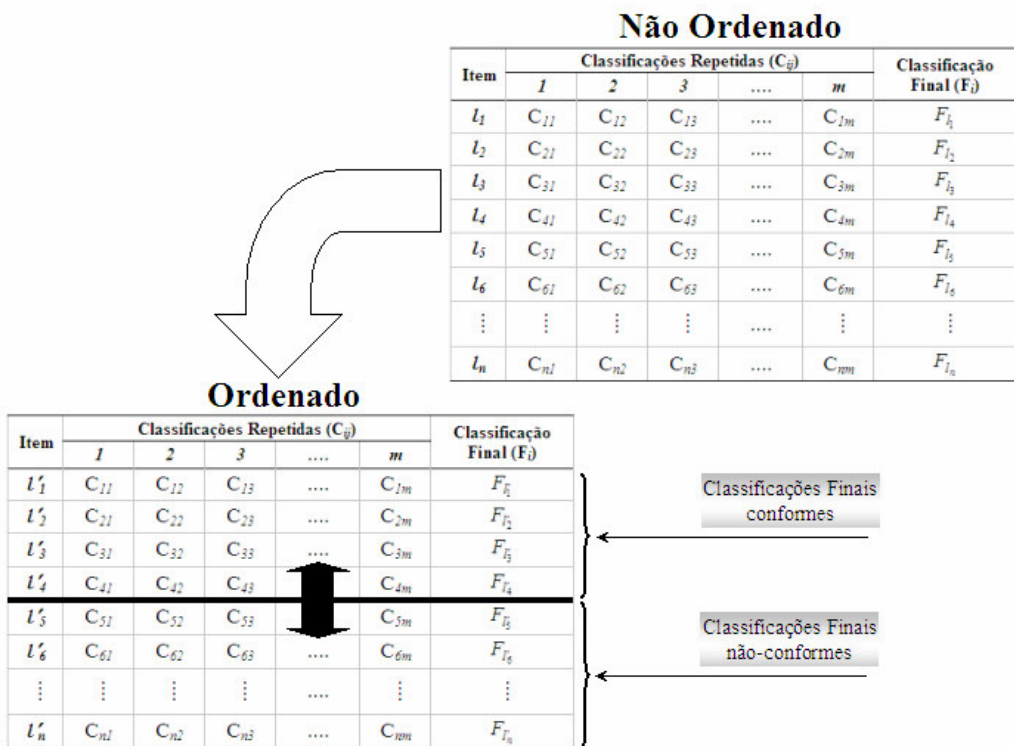


FIGURA 4.2: Ilustração do processo de ordenação da amostra pela coluna de classificação final

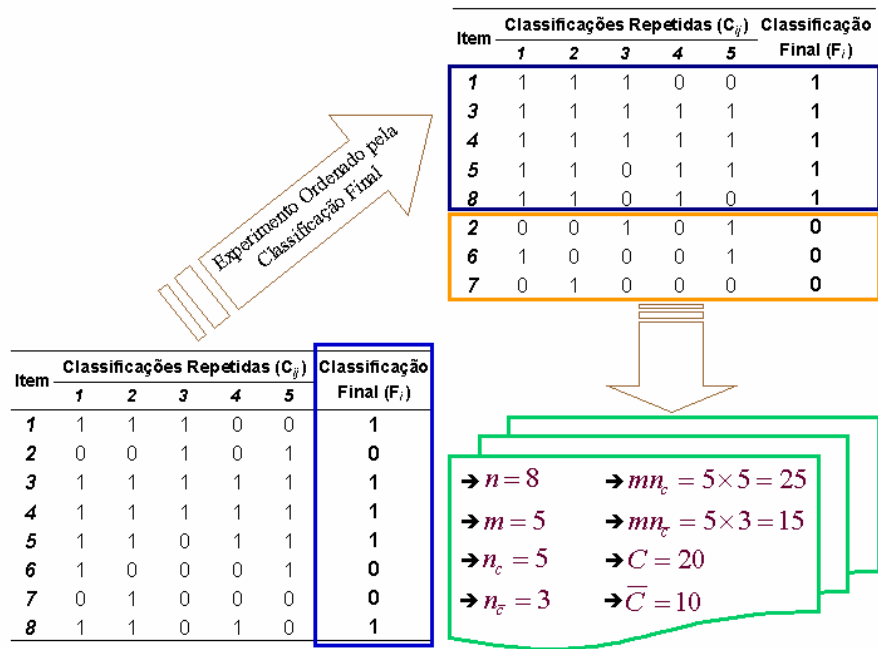


FIGURA 4.3: Ilustração de um exemplo fictício de classificação de itens

Então, com essa nova estrutura de dados, a amostra foi dividida em dois blocos: um deles contemplando apenas as classificações finais conformes e o outro indicando as classificações finais

não-conformes. Assim, considere $n_c = \sum_{i=1}^n F_i$ como sendo o número de itens classificados como

conforme e C o número total de classificações repetidas consideradas como conformes nas mn_c

classificações realizadas. Analogamente, seja $n_{\bar{c}} = n - n_c$ o número de itens classificados como não-

conformes e \bar{C} o número total de classificações repetidas consideradas como não-conformes observadas

nas $mn_{\bar{c}}$ classificações efetuadas (FIGURA 4.4).

Ordenado

Item	Classificações Repetidas (C_i)					Classificação Final (F_i)	
	1	2	3	...	m		
l'_1	C					$m \times n_C$	n_C
l'_2							
l'_3							
l'_4							
l'_5	\bar{C}					$m \times n_{\bar{C}}$	$n_{\bar{C}}$
l'_6							
\vdots							
l'_n							
l'_n							

← } Classificações Finais conformes

← } Classificações Finais não-conformes

FIGURA 4.4: Localização espacial das medidas amostrais calculadas

Utilizando um ponto de corte $m/2$ como critério para a classificação final de um determinado item como conforme, cada item que resultou numa classificação final conforme, após a realização das m inspeções, apresentou obrigatoriamente, $\left\lfloor \frac{m}{2} + 1 \right\rfloor$ classificações conformes, no mínimo, e podendo ser observado até todas as mn_C classificações conformes. Assim, considerando que foram observadas n_C classificações finais conformes, conclui-se que os valores possíveis de C estão limitados por $\left\lfloor \frac{m}{2} + 1 \right\rfloor n_C$ e mn_C , ou seja, $\left\lfloor \frac{m}{2} + 1 \right\rfloor n_C \leq C \leq mn_C$.

De forma semelhante, se para um determinado item houve uma classificação final não-conforme, após a realização das m inspeções, então significa que ocorreram, no mínimo, $\left\lfloor \frac{m}{2} + 1 \right\rfloor n_{\bar{C}}$ classificações individuais não-conformes obrigatoriamente, podendo, ainda, terem sido observadas, num limite superior, todas as classificação individuais não-conformes. Portanto, considerando que foram observadas $n_{\bar{C}}$ classificações finais não-conformes, conclui-se que o total de classificações

individuais não-conformes (\bar{C}), dentre as $mn_{\bar{C}}$ classificações realizadas, está limitado por $\left\lfloor \frac{m}{2} + 1 \right\rfloor n_{\bar{C}}$

e $mn_{\bar{C}}$, isto é, $\left\lfloor \frac{m}{2} + 1 \right\rfloor n_{\bar{C}} \leq \bar{C} \leq mn_{\bar{C}}$.

Além disso, se $n_C = 0$ ($n_{\bar{C}} = n$), isso implica que $0 \leq C \leq 0$ ($C = 0$) e que

$\left\lfloor \frac{m}{2} + 1 \right\rfloor n \leq \bar{C} \leq mn$. Da mesma forma, se $n_C = n$ ($n_{\bar{C}} = 0$), isso implica que $\left\lfloor \frac{m}{2} + 1 \right\rfloor n \leq C \leq mn$ e que

$0 \leq \bar{C} \leq 0$ ($\bar{C} = 0$).

Portanto, o interesse principal volta-se para responder a seguinte pergunta: qual é a probabilidade de observarmos n_C itens conformes, $n_{\bar{C}}$ itens não-conformes, C classificações conformes e \bar{C} classificações não-conformes? Tal probabilidade poderá ser expressa como a função de verossimilhança proposta nesta tese.

$$P(C = c, \bar{C} = \bar{c}, n_C = k) \tag{4.2}$$

Note que, uma vez observado n_C , o resultado de $n_{\bar{C}}$ é determinístico e igual a $n_{\bar{C}} = n - n_C$, não precisando, portanto, ser considerado no cálculo da probabilidade conjunta. Além disso, a expressão (4.2) pode ser reescrita como segue.

$$P(C = c, \bar{C} = \bar{c}, n_C = k) = P(C = c, \bar{C} = \bar{c} | n_C = k) P(n_C = k) \tag{4.3}$$

Observa-se que C e \bar{C} , condicionados a n_C , são independentes (condicionalmente independentes), possibilitando, portanto, a fatoração da expressão 4.3, como é mostrada a seguir:

$$\begin{aligned}
P(C = c, \bar{C} = \bar{c}, n_C = k) \\
&= P(C = c | n_C = k) P(\bar{C} = \bar{c} | n_C = k) P(n_C = k) \\
&= P(C = c | n_C = k) P(\bar{C} = \bar{c} | n_C = n - k) P(n_C = n - k)
\end{aligned}
\tag{4.4}$$

sendo

$$P(n_C = k) = \binom{n}{k} [P(F_i = 1)]^k [1 - P(F_i = 1)]^{n-k}
\tag{4.5}$$

Na expressão (4.5) $P(F_i = 1)$ representa a probabilidade de julgar um item conforme após m classificações a partir do critério da maioria ($m/2$), conforme é explicitada na expressão (3.4).

Para o cálculo de $P(C = c | n_C = k)$, considere apenas o grupo de classificações finais conformes (FIGURA 4.5). Observa-se que t dessas k classificações finais conformes podem ter sido provenientes de itens realmente conformes e $k-t$ de itens realmente não-conformes. Para o primeiro subconjunto existirão mt classificações possíveis e $m(k-t)$ classificações para o último subconjunto.

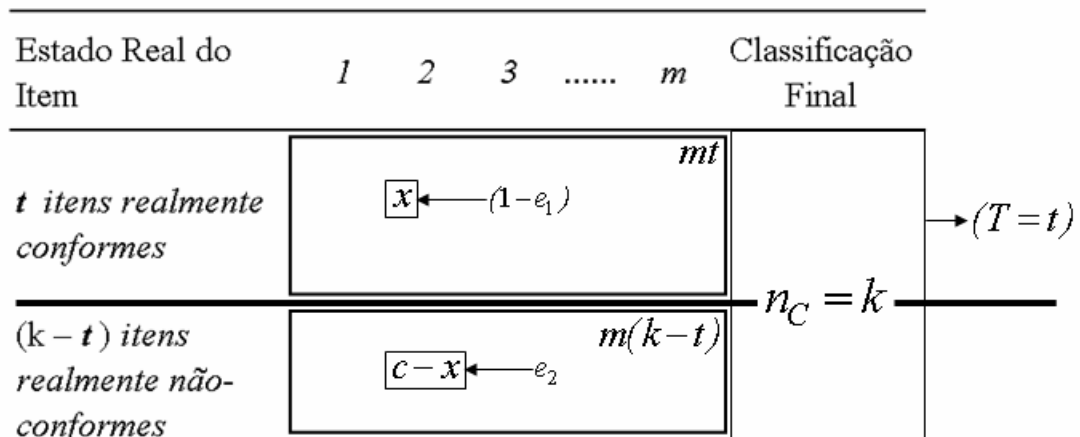


FIGURA 4.5: Subdivisão do grupo de classificações finais conformes

A probabilidade de se observar exatamente $C=c$ classificações conformes depende do número de itens realmente conformes (T), $0 \leq T \leq k$, dentre os n_C itens que tiveram classificação final conformes. Assim, $P(T = t)$ pode ser calculado como

$$P(T = t) = \binom{n_C}{t} p_C^t (1 - p_C)^{n_C - t} \quad (4.6)$$

Em que p_C indica a probabilidade de um item ser realmente conforme ($E_i = 1$), sabendo que ele foi classificado como conforme após as m classificações repetidas ($F_i = 1$), ou seja, $p_C = P(E_i = 1 | F_i = 1)$. Em termos matemáticos, o cálculo de p_C é realizado, por meio do Teorema de Bayes, como é mostrado a seguir.

$$\begin{aligned} p_C &= P(E_i = 1 | F_i = 1) \\ &= \frac{p[P(F_i = 1 | E_i = 1)]}{p[P(F_i = 1 | E_i = 1)] + (1 - p)[P(F_i = 1 | E_i = 0)]} \end{aligned} \quad (4.7)$$

Então, quanto maior for o valor de T , maior será a chance de se obter um valor de c mais alto. Além disso, para que haja exatamente $C=c$ classificações conformes, x classificações conformes deverão vir do subgrupo que contém mt classificações que sofre a influência do e_I ; e $c-x$ classificações conformes deverão vir do subgrupo que contém $m(k-t)$ classificações que é influenciado pela atuação do erro tipo II.

Portanto, a probabilidade de observar a interseção dos eventos $\{x\}$, $\{c-x\}$ e $T=t$ é dado por

$$P(C = c \mid n_C = k)$$

$$= \frac{\sum_{t=0}^k \sum_{x=M_0}^{M_1} \left[\binom{mt}{x} (1-e_1)^x e_1^{mt-x} \binom{m(k-t)}{c-x} e_2^{c-x} (1-e_2)^{m(k-t)-(c-x)} P(T=t) \right]}{\sum_{v=\lfloor \frac{m}{2} \rfloor}^{mk} \sum_{t=0}^k \sum_{x=M_0}^{M_1} \left[\binom{mt}{x} (1-e_1)^x e_1^{mt-x} \binom{m(k-t)}{v-x} e_2^{v-x} (1-e_2)^{m(k-t)-(v-x)} P(T=t) \right]} \quad (4.8)$$

Por definição, na combinatória $\binom{a}{b}$, sabe-se que $b \leq a$ ($a, b \in \mathbb{Z}^+$). Assim, do termo combinatório $\binom{mt}{x}$, tem-se que $0 \leq x \leq mt$. De forma similar, do termo combinatório $\binom{m(k-t)}{c-x}$ pode-se escrever que $0 \leq c-x \leq m(k-t)$ implicando em $c - m(k-t) \leq x \leq c$. Resumindo os dois resultados é possível afirmar que x tem de estar na interseção desses dois conjuntos, ou seja, $\max(0; c - m(k-t)) \leq x \leq \min(c; mt)$. Se $M_0 = \max(0; c - m(k-t))$ e $M_1 = \min(c; mt)$, então $M_0 \leq x \leq M_1$, como mostra a FIGURA 4.6.

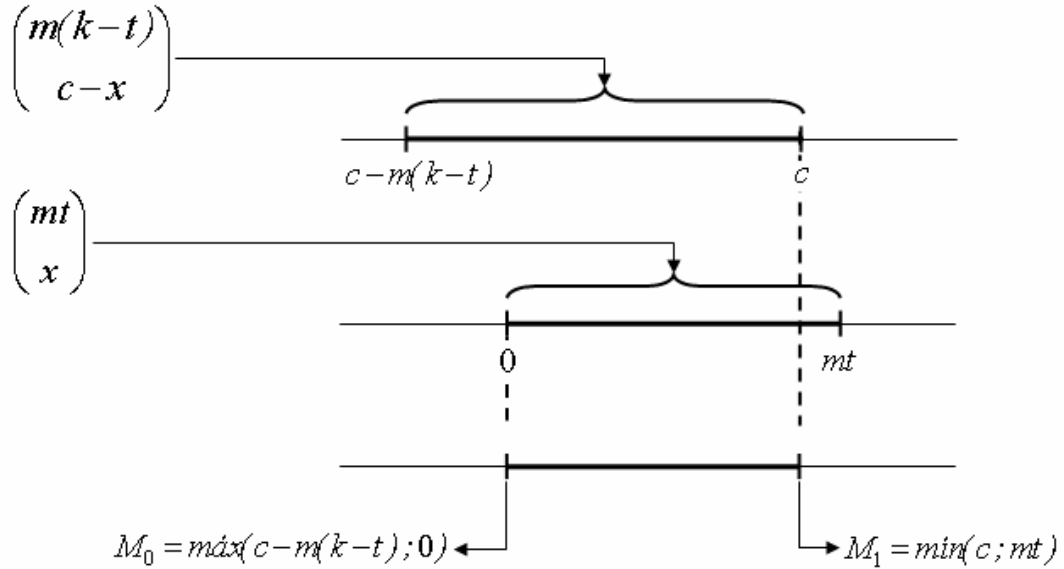


FIGURA 4.6: Obtenção dos limites M_0 e M_1

Adicionalmente, a variável aleatória C , condicionada ao conjunto $n_C = k$, reduz seu espaço amostral e, por isso, poderá assumir apenas valores que variam entre $\left\lfloor \frac{m}{2} + 1 \right\rfloor k$ e mk . Logo, a probabilidade de observar um determinado c deverá ser avaliada diante de todas as possibilidades de C estar compreendida entre $\left\lfloor \frac{m}{2} + 1 \right\rfloor k$ e mk , o que explica o denominador da expressão (4.8).

De forma similar, toda a fundamentação já apresentada pode ser empregada para explicar a probabilidade de serem observadas exatamente \bar{c} classificações não-conformes, sabendo que aconteceram $(n-k)$ classificações finais não-conformes, isto é, $P(\bar{C} = \bar{c} | n_{\bar{C}} = n - k)$. Essa probabilidade é calculada da seguinte forma

$$\begin{aligned}
 & P(\bar{C} = \bar{c} | n_{\bar{C}} = n - k) \\
 &= \frac{\sum_{z=0}^{n-k} \sum_{x=M_0}^{M_1} \left[\binom{mz}{x} (1-e_2)^x e_2^{mz-x} \binom{m(n-k-z)}{\bar{c}-x} e_1^{\bar{c}-x} (1-e_1)^{m(n-k-z)-(\bar{c}-x)} P(Z=z) \right]}{\sum_{v=\left\lfloor \frac{m}{2} + 1 \right\rfloor k}^{mk} \sum_{z=0}^k \sum_{x=M_0}^{M_1} \left[\binom{mz}{x} (1-e_2)^x e_2^{mz-x} \binom{m(n-k-z)}{v-x} e_1^{v-x} (1-e_1)^{m(n-k-z)-(v-x)} P(Z=z) \right]} \quad (4.9)
 \end{aligned}$$

em que $M_0 = \max(0; \bar{c} - m(n-k-z))$ e $M_1 = \min(\bar{c}; mz)$

Por fim, define-se variável aleatória Z como o número de itens realmente não-conformes dentre os $n_{\bar{C}}$ itens que tiveram classificação final não-conforme, de modo que $P(Z=z)$ pode ser obtido por meio da fórmula

$$P(Z = z) = \binom{n_{\bar{C}}}{z} p_{\bar{C}}^z (1 - p_{\bar{C}})^{n_{\bar{C}} - z} \quad (4.10)$$

Sendo que $p_{\bar{C}}$ representa a probabilidade de um item ser realmente não-conforme ($E_i = 0$), sabendo que sua classificação final foi não-conforme após as m classificações repetidas ($F_i = 0$), isto é, $p_{\bar{C}} = P(E_i = 0 | F_i = 0)$. O cálculo de $p_{\bar{C}}$ pode ser efetuado como

$$\begin{aligned} p_{\bar{C}} &= P(E_i = 0 | F_i = 0) \\ &= \frac{(1 - p)[P(F_i = 0 | E_i = 0)]}{p[P(F_i = 0 | E_i = 1)] + (1 - p)[P(F_i = 0 | E_i = 0)]} \end{aligned} \quad (4.11)$$

Desse modo, o produto das expressões 4.5, 4.8 e 4.9 determina a função de verossimilhança proposta aqui neste trabalho, denominada de função de verossimilhança ordenada. Essa função de verossimilhança apresenta um nível de dificuldade computacional superior às propostas discutidas por outros autores, como Lindsey (1985), Evans *et al.* (1996) e Pires (2006).

Na seção 4.3.1 será discutida a obtenção das distribuições conjunta e marginal *a posteriori*, em que se ressaltará, especialmente, o uso dos principais métodos numéricos para a sua obtenção.

4.3.1. DENSIDADES CONJUNTA E MARGINAL A POSTERIORI

Considere uma distribuição *a priori* conjunta de (p, e_1, e_2) dada por:

$$f(p, e_1, e_2) = f_{\beta}(p | \alpha, \beta) f_{\beta}(e_1 | \alpha_1, \beta_1) f_{\beta}(e_2 | \alpha_2, \beta_2) \quad (4.12)$$

em que $f_{\beta}(a|b,c)$ denota uma função densidade Beta para uma variável aleatória a com parâmetros b e c .

A FIGURA 4.7 mostra cinco possíveis distribuições betas que poderiam ser utilizadas para modelar os erros de classificação. A distribuição Beta(2,50), por exemplo, seria um modelo razoável para descrever erros de classificação que estivessem limitados a uma margem de 5%. Já a distribuição Beta(1,2) seria o modelo menos adequado, no sentido de que forneceria erros baixos e altos praticamente com as mesmas chances, ou seja, a distribuição Beta(1,2) representaria uma espécie de distribuição de probabilidade “pouco informativa” para os erros de classificação.

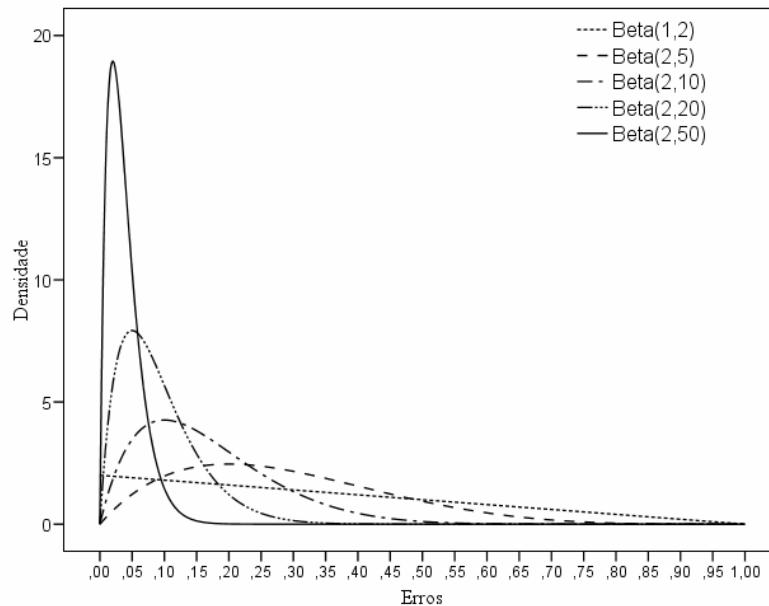


FIGURA 4.7: Modelagem dos erros de classificação por meio da distribuição beta

A distribuição Beta tem sido amplamente empregada em modelos bayesianos para descrever a informação proveniente de problemas envolvendo uma proporção (GUPTA; NADARAJAH, 2004). Assim, a função densidade conjunta *a posteriori* de (p, e_1, e_2) é obtida multiplicando a densidade *a priori* conjunta (4.15) pela função de verossimilhança (4.4) e normalizando-a conforme é requerido pelo Teorema de Bayes (WINKLER, 2003; GELMAN, 2004). Integrando com respeito a e_1 e e_2 , encontra-se a função densidade *a posteriori* marginal para o parâmetro p .

Apesar da aparente simplicidade da expressão 4.12, a maior dificuldade está relacionada à obtenção da distribuição *a posteriori* conjunta do vetor (p, e_1, e_2) e a respectiva densidade marginal *a posteriori* de p . Alguns autores como Rahme *et al.* (2000) e Stamey *et al.* (2004) sugeriram utilizar as cadeias de Markov (MCMC methods), enquanto Rubin (1988) empregou as técnicas de amostragem e reamostragem por importância ou bootstrap bayesiano ponderado e afirmou que essas estratégias de simulação podem também apresentar resultados satisfatórios. Evans, Guttman e Olkin (1993) notaram que o amostrador de Gibbs pode apresentar dificuldades quando a distribuição *a posteriori* é multimodal, situação possível no presente trabalho caso seja utilizada uma distribuição *a priori* não informativa para os parâmetros (p, e_1, e_2) . Por tudo isso, decidiu-se empregar nesse trabalho os métodos de integração numérica. Existem na literatura várias técnicas de integração numérica, entre elas a técnica do trapézio, a da quadratura gaussiana e a que utiliza métodos de simulação de Monte Carlo. Para mais detalhes, consultar Acton (1990).

Como já citado, o interesse é a obtenção da densidade marginal *a posteriori* de p , que representa a integral dupla da densidade conjunta *a posteriori* (4.13) sobre os limites de integração de e_1 e e_2 , ou seja

$$f(p) = \int_0^1 \int_0^1 f(p, e_1, e_2 | x) de_1 de_2 \quad (4.13)$$

em que $f(p, e_1, e_2 | x)$ é a densidade conjunta *a posteriori* de p .

Mas como a forma de $f(p)$ é desconhecida, resolveu-se utilizar o método de integração numérica via a técnica de simulação de Monte Carlo. A idéia principal do método de integração de Monte Carlo é aproximar a área sob uma curva $f(x)$ por meio da área de um retângulo. A FIGURA 4.8 ilustra a aproximação de uma integral I sob a curva $f(x)$ pela área Q , que representa o método de integração de Monte Carlo para o caso unidimensional.

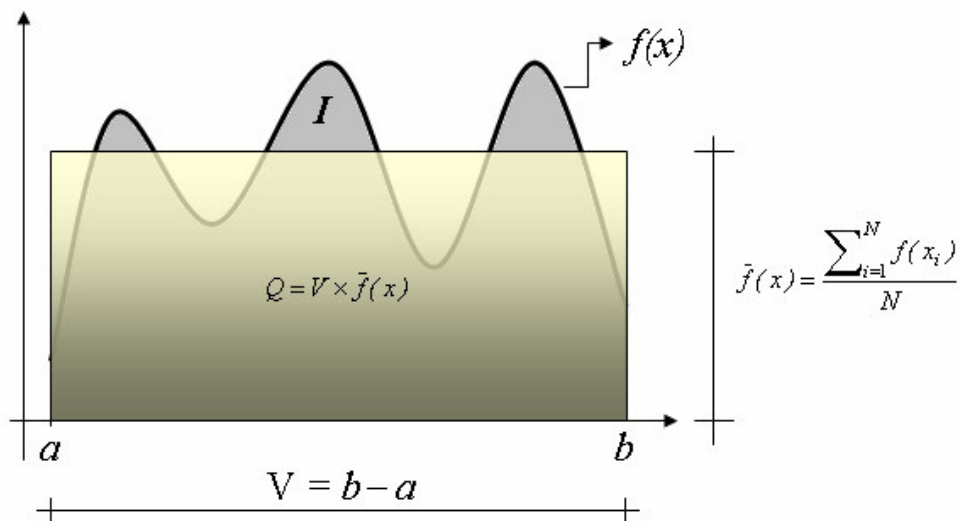


FIGURA 4.8: Ilustração do método de integração de Monte Carlo para o caso unidimensional

No caso r -dimensional, considere a integral de uma função $f(u_1, u_2, \dots, u_r)$ dependendo das variáveis u_1, u_2, \dots, u_r com domínios $a_1 \leq u_1 \leq b_1, a_2 \leq u_2 \leq b_2, \dots, a_r \leq u_r \leq b_r$, respectivamente e sobre o hipercubo de volume $V = (b_1 - a_1) \times (b_2 - a_2) \times \dots \times (b_r - a_r)$. Seja $s = (u_1, u_2, \dots, u_r)$ um ponto nesse hipercubo e $f(s) = f(u_1, u_2, \dots, u_r)$ a função avaliada nesse ponto s . Se é extraída uma amostra aleatória de tamanho N de vetores s , então, a aproximação Monte Carlo (Q) da integral I é dada por

$$I = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_r}^{b_r} f(s) du_1 du_2 \cdots du_r \approx Q = \frac{V}{N} \sum_{i=1}^N f(s_i) \quad (4.14)$$

A Lei Forte dos Grandes Números e o Teorema Central do Limite garantem que Q converge quase certamente para I e possui distribuição Normal.

Mas no caso específico desta tese, o interesse não é calcular a área sob a curva $f(p)$, mas sim a sua função densidade de probabilidade (FIGURA 4.9). Por isso, foi desenvolvido o seguinte procedimento:

PASSO 1: Dividir o espaço paramétrico de p em uma malha bem fina de pontos, sendo p^* um ponto genérico desse intervalo;

PASSO 2: Para o ponto p^* gerar uma amostra aleatória de tamanho N do vetor $s = (e_1, e_2)$ proveniente de uma distribuição Beta(α_i, β_i), $i=1, 2$, dos parâmetros e_1 e e_2 ;

PASSO 3: Aplicar o vetor s na densidade conjunta *a posteriori* de p , isto é, $f(p^*, s | x)$, produzindo um arquivo com N medições;

PASSO 4: Obter a média do arquivo construído no PASSO 3 e calculou-se Q como mostrado na expressão (4.14), sendo $V=1$ o volume do hipercubo sobre o espaço paramétrico de (e_1, e_2) ;

PASSO 5: Guardar os valores de p^* e de Q num arquivo, para todos os valores de p^* ;

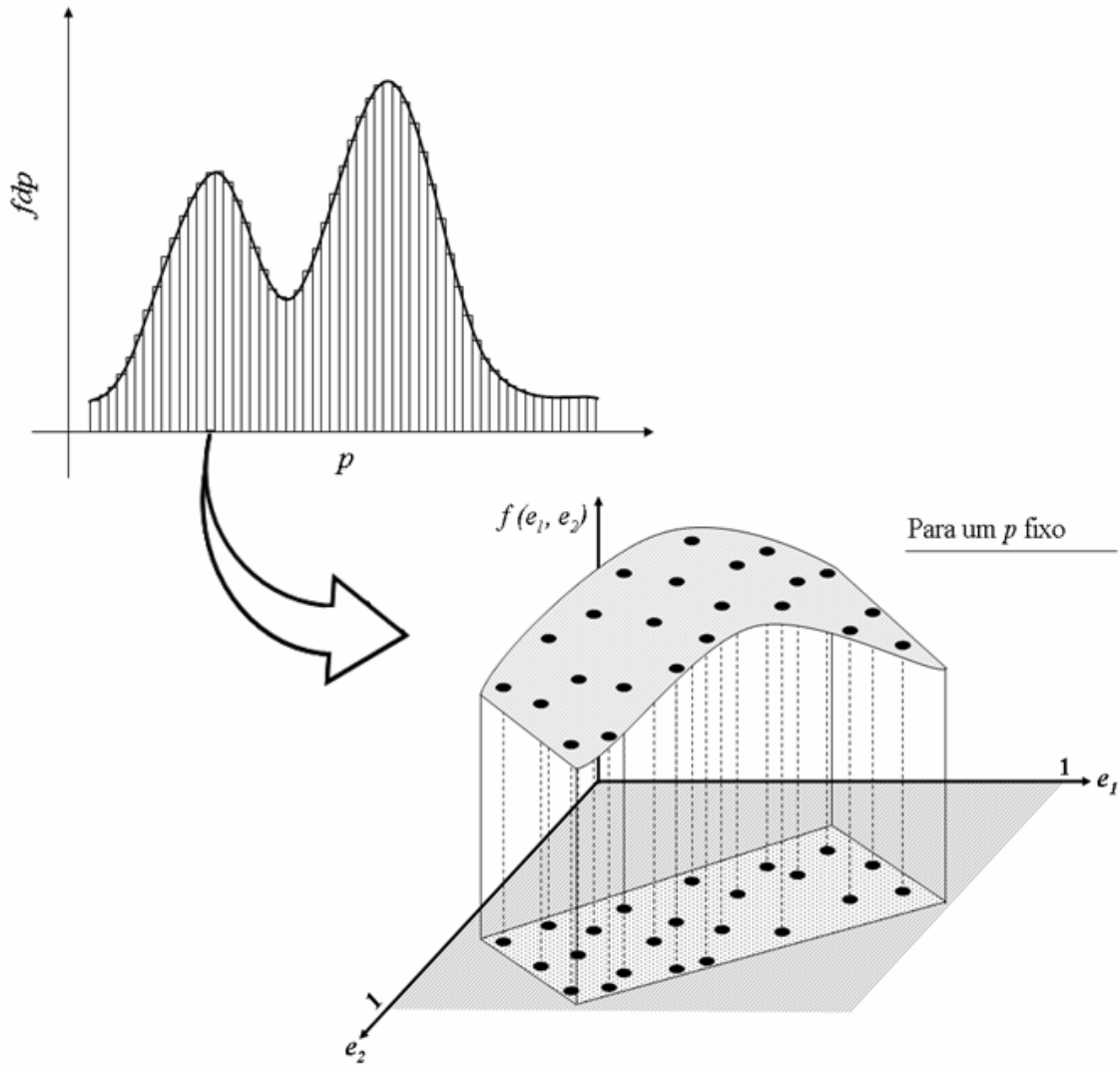


FIGURA 4.9: Obtenção da densidade marginal *a posteriori* de p via Integração de Monte Carlo

Como esse mesmo raciocínio, é possível também implementar a construção das distribuições *a posteriori* marginais de e_1 e e_2 . Nessa tese, sem perda de generalidade, referir-se-á à distribuição *a posteriori* marginal de p como simplesmente distribuição *a posteriori* de p . E o algoritmo do processo completo de simulação encontra-se descrito a seguir.

ALGORITMO 1: OBTENÇÃO DA DENSIDADE A POSTERIORI DE p

Entrada: n (número de itens avaliados).
 m (número de classificações repetidas).
 p (proporção de itens conformes).
 e_1 (probabilidade de classificar um item conforme erroneamente como não-conforme).
 e_2 (probabilidade de classificar um item não-conforme erroneamente como conforme).
 $Beta(\lambda, \tau)$ (distribuição *a priori* para os erros de classificação com parâmetros λ e τ).

Saída: distribuição marginal *a posteriori* de p e suas medidas estatísticas (desenvolvido na plataforma do *Statistical Analysis System – SAS*, versão 9.1.3)

PASSO 1: Geração do Experimento

Simular o experimento a partir da definição dos parâmetros n, m, p, e_1 e e_2 ;

PASSO 2: Obtenção das quantidades analíticas

A partir do experimento obtém-se as quantidades $n_C, n_{\bar{C}}, c$ e \bar{c} ;

PASSO 3: Definição da Malha de Pontos do Espaço Paramétrico Investigado

Define-se a malha de pontos e o tamanho da janela (h) sobre a qual será construída a Densidade *a posteriori* de p ;

PASSO 4: Método de Aproximação de Integral de Monte Carlo

Constrói-se um *looping* sobre o espaço paramétrico de p e, a partir da técnica de simulação de Monte Carlo, estima-se a integral dupla sobre (e_1, e_2) , guardando-se, em seguida, o valor da função densidade *a posteriori* de p e o próprio ponto p ;

PASSO 5: Obtenção das Medidas Estatísticas

A partir da matriz contendo os valores do parâmetro p e de sua respectiva função densidade *a posteriori*, estimam-se as medidas de qualidade do processo de simulação, a citar: moda, desvio padrão e tamanhos dos intervalos de credibilidade para o parâmetro p ;

Um outro interesse relacionado ao estudo da distribuição marginal *a posteriori* de p é a determinação do seu intervalo de credibilidade. Por definição, G é um intervalo de credibilidade de $100(1-\alpha)\%$ para p se $P(p \in G) = 1 - \alpha$ (HYNDMAN, 1996), $0 < \alpha < 1$. Note que a definição expressa de forma probabilística a pertinência ou não de p ao intervalo. Isso significa que, para $G=[a,b]$,

quanto menor for a amplitude do intervalo ($w=b-a$), mais concentrada será a distribuição do parâmetro, ou seja, a amplitude do intervalo informa sobre a dispersão de p . Esse assunto será tratado na seção 4.4.

4.4. INTERVALO DE CREDIBILIDADE DE MÁXIMA DENSIDADE A POSTERIORI

A obtenção da região G pode não ser uma tarefa fácil de ser realizada, pois existe uma infinidade de intervalos que atendem à especificação mencionada anteriormente. Por isso, nos últimos anos, vem sendo dada uma atenção maior a uma classe de intervalos de credibilidade que se baseia na determinação daquelas áreas que apresentam uma maior densidade *a posteriori*. Assim, diz-se que um intervalo de credibilidade G de $100(1-\alpha)\%$ para p é de máxima densidade *a posteriori* (MDP) se $G = \{p \in \Theta : f(p|x) \geq k(\alpha)\}$ em que $k(\alpha)$ é a maior constante, tal que $P(p \in G) = 1 - \alpha$, $0 < \alpha < 1$. Em palavras mais simples, todos os pontos pertencentes a um intervalo MDP terão densidade maior do que qualquer outro não pertencente ao intervalo.

Para distribuições unimodais e simétricas, os intervalos de credibilidade tradicional e o MDP fornecem limites idênticos. Mas havendo assimetria, os dois critérios podem produzir intervalos diferentes e essa divergência se acentua à medida que aumenta o grau de assimetria. Se, por outro lado, a distribuição *a posteriori* for multimodal, o conjunto G pode ser constituído de vários sub-intervalos em torno das modas mais destacadas. Nesse caso, os intervalos MDP levam vantagem sobre os intervalos de credibilidade tradicionais. Para mais detalhes, consulte Berger (1985) e Box e Tiao (1992). Nesta tese, as amplitudes dos intervalos MDP das distribuições *a posteriori* de p dos modelos FVO e FVC serão utilizadas para compará-los.

4.5. COMPARAÇÃO DOS MODELOS VIA FATOR DE BAYES

A ciência é fundamentada em informações a respeito de uma determinada quantidade de interesse θ e a toda informação está associada ao grau de incerteza sobre esse parâmetro. Assim, um dos objetivos da inferência bayesiana é minimizar esse grau de incerteza sobre θ . Porém, a incerteza a respeito de θ pode assumir diferentes graus, sendo, modelado, geralmente, por uma distribuição de probabilidade.

O grau de incerteza, ou a quantidade de informação, em relação a θ , é representado, probabilisticamente, por meio de uma função densidade $h(\theta)$. A informação sobre θ pode ainda ser aumentada observando-se uma quantidade aleatória X relacionada com θ , na qual a distribuição amostral $f(x|\theta)$ define essa relação. Assim, uma vez observada $X=x$, a quantidade de informação sobre θ aumenta e o teorema de Bayes é a regra de atualização utilizada para mensurar esse aumento de informação, ou seja,

$$f(\theta|x) = \frac{f(x;\theta)}{f(x)} = \frac{f(x|\theta)h(\theta)}{f(x)} = \frac{f(x|\theta)h(\theta)}{\int f(x|\theta)h(\theta)d\theta} \quad (4.15)$$

Observe, na equação 4.15, que $f(x)$ desempenha o papel de constante normalizadora de $f(\theta|x)$, pois não depende de θ . Além disso, para um valor fixo de x , a função $f(x|\theta)$ fornece a verossimilhança de cada um dos valores de θ , ou seja, $f(x|\theta)=L(x;\theta)$, enquanto que $h(\theta)$ designa a distribuição *a priori* de θ . Essas duas fontes de informação, *a priori* e a verossimilhança, são combinadas formando a distribuição *a posteriori* de θ , nominalmente conhecida como $f(\theta|x)$. Portanto, a forma usual do Teorema de Bayes é

$$f(\theta|x) \propto L(x;\theta) \times h(\theta) \quad (4.16)$$

É importante ressaltar que essa forma simplificada do Teorema de Bayes (expressão 4.15) é utilizada na resolução de problemas que envolvam estimação de parâmetros, já que o denominador é apenas uma constante normalizadora. Porém, em outras situações, como a de seleção de modelos, esse termo tem um papel crucial.

A abordagem bayesiana aplicada a testes de hipóteses foi desenvolvida por Jeffreys (1935), apesar de denominá-la de testes de significância. Sob a ótica de Jeffreys, a técnica consistia em comparar previsões feitas por duas teorias científicas, em que modelos estatísticos eram introduzidos para representar a probabilidade dos dados de acordo com cada uma das teorias e o teorema de Bayes era usado para calcular a probabilidade *a posteriori* de uma das teorias ser correta. Mais especificamente, a quantidade de informação utilizada para decidir sobre qual das teorias estaria correta ficou conhecida, atualmente, como fator de Bayes. Kass e Raftery (1995) oferecem uma discussão detalhada sobre as definições e aplicações envolvendo o fator de Bayes.

Assim, o conhecimento *a posteriori* do comportamento de um determinado parâmetro θ depende, basicamente, da informação conjunta contida na distribuição *a priori* e na verossimilhança proveniente da observação de um conjunto de dados. Essas são razões suficientemente fortes para entender que existe uma família de modelos que pode ser utilizada para descrever os dados e os parâmetros. Assim, a classe de modelos descrita abaixo

$$\mathfrak{S} = \{f_i(x|\theta_i), h_i(\theta_i), i \in I, \theta_i \in \Theta\}$$

contempla o caso mais geral em que as distribuições amostrais são diferentemente parametrizadas, sendo I um conjunto, que pode ser discreto ou contínuo, que contém o modelo verdadeiro, mas desconhecido.

Vários são os casos especiais que se pode considerar dependendo de cada situação concreta. Um deles é a classe que contempla um ou mais modelos para os dados dotados do mesmo parâmetro e da mesma distribuição *a priori*, variando apenas a sua forma distribucional, ou seja

$$\mathfrak{S}_1 = \{f_i(x|\theta), h_i(\theta), i \in I, \theta \in \Theta\}$$

Uma outra classe importante é aquela que contempla os modelos preditivos *a priori*, isto é, $\mathcal{M} = \{M_i, i \in I, \theta \in \Theta\}$, cujas distribuições são definidas por

$$p(x|M_i) = \int f_i(x|\theta_i)h_i(\theta_i)d\theta_i \quad (4.17)$$

No caso específico de serem considerados a mesma distribuição *a priori* e o mesmo parâmetro na modelagem dos dados, a expressão 4.17, torna-se

$$p(x|M_i) = \int f_i(x|\theta)h(\theta)d\theta \quad (4.18)$$

em que a distribuição preditiva global corresponde à

$$p(x) = \int_I p(x|M_i)p(M_i)dM_i \quad (4.19)$$

sendo que $p(M_i)$ é a probabilidade da *a priori* de M_i ser o modelo verdadeiro que, atualizado em face aos dados x , obtém-se

$$p(M_i | x) = \frac{p(x | M_i) p(M_i)}{p(x)}, \quad i \in I \quad (4.20)$$

Desse modo, o fator de Bayes que compara os modelos M_1 e M_2 é formalmente definido como sendo a razão entre as distribuições amostrais condicionadas nos dois modelos sob avaliação. A demonstração é mostrada a seguir.

$$B_{12} = \frac{p(x | M_1)}{p(x | M_2)} = \frac{\frac{p(M_1 | x) p(x)}{p(M_1)}}{\frac{p(M_2 | x) p(x)}{p(M_2)}} = \frac{p(M_1 | x)}{p(M_2 | x)} \times \frac{p(M_2)}{p(M_1)} = \quad (4.21)$$

Como visto na equação 4.21, percebe-se que o fator de Bayes em favor do modelo M_1 é a razão entre as respectivas distribuições preditivas sob M_1 e M_2 (4.18). Assim, um fator de Bayes maior que 1 indicaria que o modelo M_1 , para uma mesma distribuição *a priori* e para um mesmo experimento x , melhor representou toda a informação probabilística *a posteriori* acerca de θ à luz dos dados observados, comparado com o modelo M_2 .

Conforme descrito na equação 4.17, verifica-se que a maior dificuldade em se calcular o fator de Bayes é a obtenção das distribuições preditivas $p(x | M_1)$ e $p(x | M_2)$ que, não necessariamente, apresentam uma forma bem definida. Por isso, para esses casos, é mais conveniente a aplicação de métodos numéricos para a aproximação dessas distribuições. Loschi e Franco (2006) apresentaram um estudo de simulação que utilizou as técnicas de amostragem por importância e da regra de Simpson para extrair informação da distribuição *a posteriori*. Além disso, Loschi e Franco (2006) empregaram simulação de Monte Carlo para comparar a influência de diferentes tamanhos amostrais e de várias especificações da distribuição *a priori* sobre os estimadores de Bayes e de máxima verossimilhança da fração de não-disjunção em anomalias cromossômicas.

Nesta tese, o fator de Bayes será utilizado para comparar o modelo que utiliza a função de verossimilhança Ordenada (FVO), intitulado de M_1 , e aquele que foi estudado por Evans *et al.* (1996), baseada na função de verossimilhança Completa (FVC), denominado aqui de M_2 . Adaptando essas denominações à equação 4.21 e aos objetivos dessa tese, o cálculo do fator de Bayes, depois de observados os dados x , isto é, m , n_C , C , $n_{\bar{C}}$ e \bar{C} , torna-se a razão das constantes de normalização dos dois modelos, ou seja

$$B_{12} = \frac{p(x|M_1)}{p(x|M_2)} = \frac{\int_0^1 \int_0^1 \int_0^1 f_1(x|p, e_1, e_2) f(p, e_1, e_2) dp de_1 de_2}{\int_0^1 \int_0^1 \int_0^1 f_2(x|p, e_1, e_2) f(p, e_1, e_2) dp de_1 de_2} \quad (4.23)$$

em que $f_1(x|p, e_1, e_2)$ e $f_2(x|p, e_1, e_2)$ são as respectivas funções de verossimilhança dos modelos proposto (4.4) e o padrão (4.1), enquanto que $f(p, e_1, e_2)$ é a distribuição conjunta *a priori* dos parâmetros p , e_1 e e_2 está descrita em 4.12.

Devido à dificuldade de estimação das respectivas distribuições preditivas $p(x|M_1)$ e $p(x|M_2)$ de cada modelo, novamente, o método de integração numérica de Monte Carlo será empregado para aproximar tais distribuições. O procedimento desenvolvido para realizar esse estudo comparativo será discutido nas próximas seções.

4.6. APLICAÇÃO A EXEMPLOS NUMÉRICOS

Nesta seção serão discutidos dois exemplos numéricos que servirão para ilustrar a aplicação da metodologia proposta nesta tese.

O primeiro exemplo foi baseado no estudo relatado por Garcia *et al.* (1997) e que também foi usado como exemplo numérico por Tang *et al.* (2004). O objetivo do estudo foi comparar a fração de resultados positivos verdadeiros (sensitividade) de dois procedimentos de detecção de alergia à penicilina. Para investigar isso, um grupo de 30 indivíduos, conhecidamente alérgicos à droga, foi selecionado, dos quais 26 apresentaram resultados positivos, segundo ambos os procedimentos. Isso indica que a probabilidade estimada de qualquer um dos procedimentos de identificar um indivíduo alérgico à penicilina é de 0,8667.




Uma pequena adaptação do problema será feita para exemplificar os procedimentos estudados. Suponha, então, que apenas um dos dois procedimentos seja aplicado a cada indivíduo m vezes e que os resultados estejam sujeitos a erros de classificação de falso-positivo (e_1) e falso-negativo (e_2).

Assim, foram simuladas $m=3, 7$ e 11 classificações repetidas independentes para cada indivíduo, e como ilustração, considere $e_1=e_2=0,05$. Além disso, foi considerado que 26 desses pacientes apresentaram respostas positivas à aplicação do teste e quatro tiveram respostas negativas. Os dados experimentais simulados contendo todas as classificações repetidas são apresentados na TABELA 4.1, sendo **0** uma classificação não-conforme e **1** uma classificação conforme.

TABELA 4.1: Dados amostrais simulados para o estudo apresentado por Garcia *et al.* (1997) para $m=3, 7$ e 11 classificações repetidas, $e_1=e_2=0,05$ e distribuições *a priori* independentes $Beta(1,2)$

para os erros de julgamento

PACIENTE	CLASSIFICAÇÕES										
	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0
5	1	0	0	1	1	1	1	1	1	1	0
6	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	0	1	1	1
8	1	1	1	0	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	0	1	0	1	1	1	1	1
11	1	1	0	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	0	1	1	1
15	1	1	1	1	1	1	0	1	1	1	1
16	1	1	0	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	1	1
19	0	1	1	0	1	1	1	1	1	1	0
20	1	1	1	1	1	1	1	1	1	1	1
21	1	1	1	1	1	1	1	0	1	1	1
22	1	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1	1
25	1	1	1	1	1	1	1	1	1	1	1
26	1	1	1	1	1	1	1	1	1	1	1
27	1	1	1	1	1	1	1	1	1	0	1
28	1	1	1	1	0	1	1	1	1	1	1
29	1	1	0	1	1	1	1	1	1	1	1
30	1	0	1	1	1	1	1	1	1	1	1

$m=3$ 
 $m=7$ 
 $m=11$ 

Teoricamente, três planilhas distintas, como àquela apresentada na TABELA 4.1, poderiam ter sido geradas para cada caso considerado ($m=3, 7, 11$) de classificações repetidas. No entanto, como estratégia de economia de tempo, geraram-se os dados somente para $m=11$ e, a partir desse, foram aproveitadas os grupos de $m=3$ e $m=7$ classificações para obter as medidas inferenciais n_C , $n_{\bar{C}}$, C e \bar{C} (TABELA 4.1) do modelo bayesiano proposto. Essa atitude não compromete em nada os resultados, pois testes simulados foram feitos e as estatísticas *a posteriori* se mostraram próximas. Esses cálculos são mostrados na TABELA 4.2 e serão utilizados para obter as distribuições *a posteriori* de p para os três grupos discutidos.

TABELA 4.2: Estatísticas inferenciais para cada grupo de classificações repetidas

Classificações Repetidas Consideradas (m)	Estatísticas da Amostra			
	n_C	$n_{\bar{C}}$	C	\bar{C}
3	25	5	70	13
7	26	4	169	25
11	26	4	267	41

Para obter as distribuições *a posteriori* da verdadeira proporção (p) de resultados positivos foram geradas 100.000 amostras de Monte Carlo. Os resultados apresentados na FIGURA 4.10 revelam a qualidade da distribuição *a posteriori* da proporção de resultados positivos obtidos pelo procedimento escolhido, uma vez que a curva apresenta um perfil unimodal, concentrada no verdadeiro valor de p e com variância inversamente proporcional ao número de classificações realizadas. Observa-se, ainda, que o aumento do número de classificações proporciona uma moda *a posteriori* cada vez mais próxima do verdadeiro valor do parâmetro p , isto é, a moda que se aproxima rapidamente de $p=0,8667$.

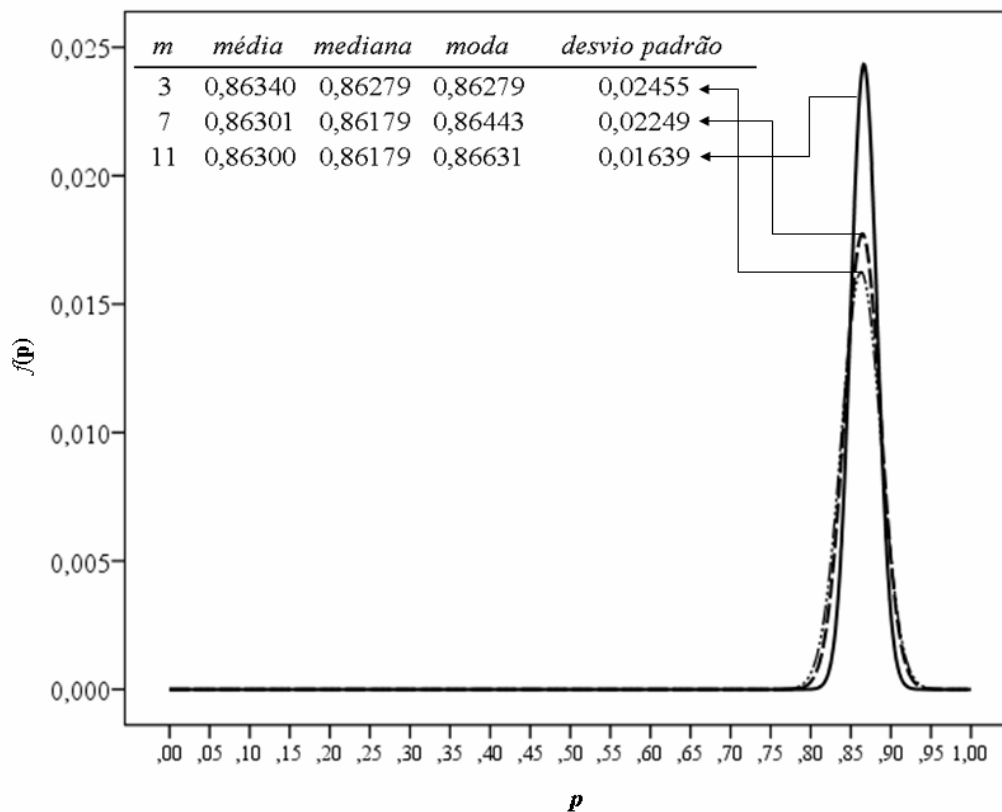


FIGURA 4.10: Distribuição *a posteriori* de p para o exemplo numérico apresentado por Garcia *et al.*

Essa convergência corrobora positivamente com o uso de classificações repetidas para situações práticas que contém erros de julgamento. Também justifica o emprego da função de verossimilhança ordenada em conjunto com técnicas de inferência bayesiana, uma vez que, à medida que o número de classificações repetidas aumenta, aumenta também a precisão da estimativa sobre a proporção de resultados positivos, pois o desvio padrão se torna cada vez menor.

O segundo exemplo é baseado no estudo apresentado por Fujisawa e Izume (2000) que estimaram o percentual de alelos de quatro antígenos (M, N, S, s) para as cidades de Hiroshima e Nagasaki. Os autores consideraram 2, 3 e 4 testes de serologia aplicados a pacientes de um programa

de saúde adulta e de crianças nascidas na primeira geração de pais sobreviventes, cujo histórico consiste em dados coletados entre 1963 e 1994.

Especificamente, para os dados referentes ao antígeno N da cidade de Hiroshima com 3 classificações repetidas, foram observados 5, 10 e 225 experimentos respectivamente com 1, 2 ou 3 respostas positivas e 117 experimentos com respostas negativas, totalizando um histórico de 357 medições. A TABELA 4.3 descreve os resultados amostrais obtidos considerando a função de verossimilhança proposta. Para esses dados, Fujisawa e Izume (2000) utilizaram um modelo de efeitos aleatórios para os erros de julgamento e obtiveram uma estimativa de 0,655 para o verdadeiro percentual de alelos contendo o antígeno N e com limites de 0,607 e 0,701 para um intervalo de credibilidade de 95%.

TABELA 4.3: Frequência Observada de Respostas Positivas (k) entre os Testes Serológicos Repetidos (m) para o Antígeno N da cidade de Hiroshima

$k(m=3) \text{ e } n=357$			
0	1	2	3
117	5	10	225

Assim, com o intuito de avaliar o desempenho do modelo proposto, foi planejado um estudo comparativo com os modelos desenvolvidos por Pires (2006) e Evans *et al.* (1996), todos com três níveis de informação para as distribuições *a priori* dos erros de classificação. A TABELA 4.4 descreve as estruturas das distribuições *a priori* para os modelos considerados.

TABELA 4.4: Distribuições *a priori* para os modelos que serão simulados

Modelo	Distribuições <i>a priori</i>	
	e_1	e_2
Modelo 1 (M1) – Proposto	Beta (1,2)	Beta (1,2)
Modelo 2 (M2) – FVC	Beta (1,2)	Beta (1,2)
Modelo 3 (M3) – Pires (2006)	Beta (1,2)	Beta (1,2)
Modelo 4 (M4) – Versão 1 de M1	Beta (2,10)	Beta (2,10)
Modelo 5 (M5) – Versão 1 de M2	Beta (2,10)	Beta (2,10)
Modelo 6 (M6) – Versão 1 de M3	Beta (2,10)	Beta (2,10)
Modelo 7 (M7) – Versão 2 de M1	Beta (2,48)	Beta (2,48)
Modelo 8 (M8) – Versão 2 de M2	Beta (2,48)	Beta (2,48)
Modelo 9 (M9) – Versão 2 de M3	Beta (2,48)	Beta (2,48)

Como já discutido, do ponto de vista prático, é desejável que os erros de julgamento sejam valores pequenos. Desse modo, deve-se tentar definir distribuições *a priori* que reflitam esse comportamento, ou seja, dando maior probabilidade aos valores mais próximos de zero desses erros. A FIGURA 4.7 ilustra bem esse sentimento apresentando cinco modelos de distribuições betas para os erros. Portanto, o efeito prático de se determinar, por exemplo, uma distribuição *a priori* Beta(2,48) para os erros e_1 e e_2 , é garantir que a probabilidade de serem selecionados erros elevados é pequena. Por isso, foram escolhidos as distribuições *a priori* Beta (1;2), Beta (2;10) e Beta (2;48) para representar os erros de classificação, com o intuito de aumentar gradativamente as concentrações das áreas com maior densidade nos valores menores de e_1 e e_2 .

Como pode ser observado na TABELA 4.4, os modelos 1, 4 e 7 compõem três versões da metodologia proposta neste trabalho contendo três níveis de informação para as distribuições *a priori* dos erros de classificação: Beta (1;2); Beta (2;10) e Beta (2;48). Da mesma forma, para os modelos abordados por Pires (2006) e pela função de verossimilhança Completa (FVC) foi

construída uma estrutura análoga. Note, ainda, que em todos os casos foi considerada uma distribuição *a priori* uniforme para a proporção de itens conformes, seguindo Evans *et al.* (1996). A FIGURA 4.10 mostra as distribuições *a priori* utilizadas no estudo comparativo.

Desse modo, as nove propostas apresentadas foram aplicadas aos dados da cidade de Hiroshima relativo ao antígeno N, considerando três testes serológicos e os resultados são apresentados na FIGURA 4.11.

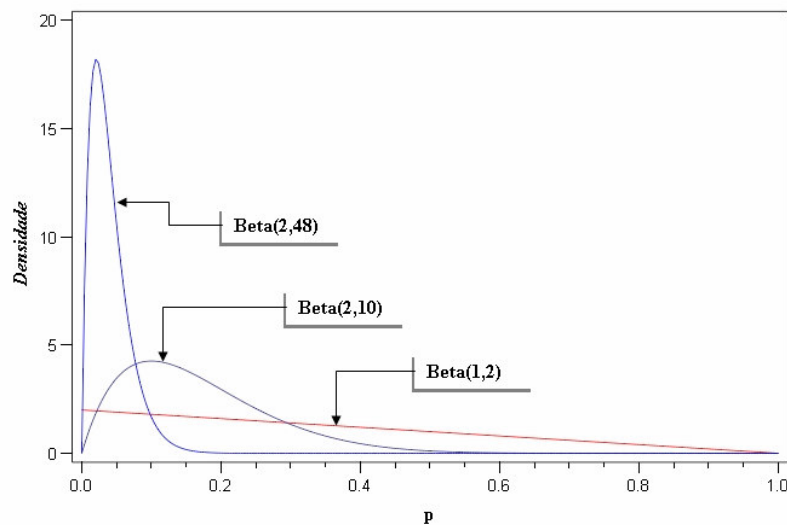


FIGURA 4.11: Distribuições *a priori* consideradas no estudo comparativo

De um modo geral, percebe-se que as três metodologias apresentaram um desempenho satisfatório, pelo menos com relação à locação de suas respectivas distribuições *a posteriori*. Além disso, observa-se, ainda, que a quantidade de informação contida nas distribuições *a priori* dos modelos contribuiu positivamente para o aumento da precisão das estimativas, ou seja, para a diminuição da variabilidade (TABELA 4.5). Esse comportamento, de certa forma, já era esperado, uma vez que, um maior conhecimento prévio sobre um determinado parâmetro aumenta também a chance de fazer previsões mais precisas sobre o mesmo.

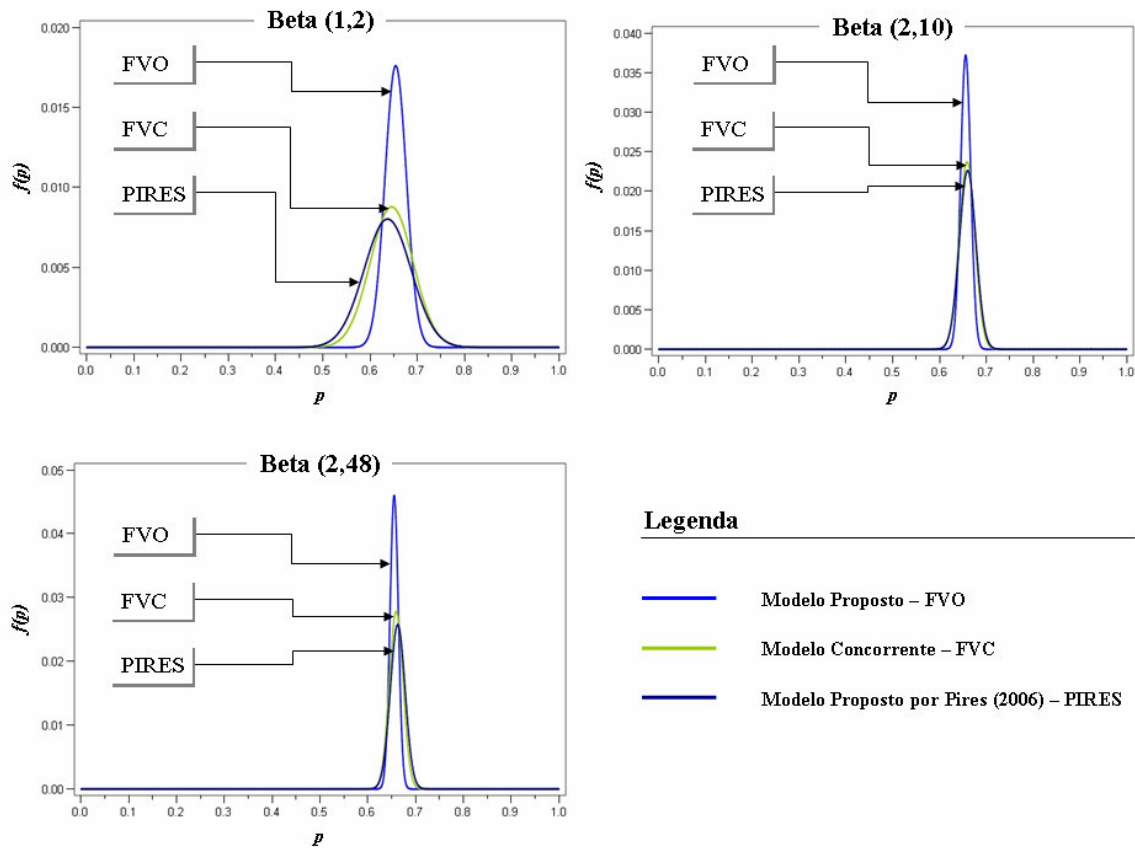


FIGURA 4.12: Distribuições *a posteriori* de p para nove propostas de modelagem aplicadas ao estudo numérico apresentado por Fujisawa e Izume (2000)

Mais especificamente, constata-se que a versão original do modelo proposto (erros com pouca informação *a priori*) se mostrou muito próxima da formulação e dos resultados descritos por Fujisawa e Izume (2000), inclusive com amplitude de intervalo de credibilidade de 0,0901, que é ligeiramente inferior ao destes.

Paralelamente, nota-se, também, que o modelo que emprega a verossimilhança completa, apesar de uma ligeira vantagem em termos de comprimento do intervalo de credibilidade, praticamente equiparou-se ao modelo estudado por Pires (2006) para todos os níveis de informação da distribuição *a priori* dos erros de classificação. Essa semelhança, em termos de desempenho, dos modelos FVC e de Pires (2006) talvez seja explicada pelo elevado número de respostas positivas

para $m=2$ e $m=3$ (TABELA 4.3). Isso contribuiu para que, na função de verossimilhança estudada por Pires (2006), houvesse muitas classificações finais conformes, melhorando, assim, a estimativa da proporção de respostas positivas. Da mesma forma, o modelo FVC também se beneficiou, pois a sua função de verossimilhança foi influenciada principalmente pela potência 225, que representa o número de respostas positivas nos três testes realizados.

Quando a metodologia sugerida nesta tese (FVO) é comparada com os modelos de Pires (2006) e Evans *et al.* (1996), os resultados indicam uma melhor performance para o modelo proposto, uma vez que apresentou menor desvio padrão *a posteriori* e menor amplitude do intervalo de credibilidade para todos os casos avaliados.

Um outro detalhe que pode ser observado é o desempenho similar entre a formulação padrão dos autores Fujisawa e Izume (2000) e a do modelo **M1** (TABELA 4.5), mostrando que é necessário somente uma distribuição *a priori* Beta(1,2) para os erros de classificação para que os resultados obtidos pela técnica proposta e os resultados de Fujisawa e Izume (2000) sejam similares. Os resultados são apresentados na TABELA 4.5.

TABELA 4.5: Medidas estatísticas para a distribuição *a posteriori* de p

Distribuição <i>a priori</i> para os erros	Modelo	Média	Mediana	Moda	Desvio Padrão	Amplitude do IC (95%)
Beta (1,2)	FVO (M1)	0,6544	0,6536	0,6546	0,0228	0,0901
	FVC (M2)	0,6458	0,6446	0,6456	0,0455	0,1802
	PIRES (M3)	0,6372	0,6366	0,6376	0,0499	0,1982
Beta (2,10)	FVO (M4)	0,6558	0,6546	0,6556	0,0110	0,0440
	FVC (M5)	0,6587	0,6576	0,6586	0,0170	0,0661
	PIRES (M6)	0,6600	0,6586	0,6596	0,0178	0,0721
Beta (2,48)	FVO (M7)	0,6552	0,6546	0,6556	0,0090	0,0340
	FVC (M8)	0,6594	0,6586	0,6596	0,0146	0,0581
	PIRES (M9)	0,6626	0,6616	0,6626	0,0157	0,0621

Apesar dos indicativos que o modelo FVO apresenta uma variabilidade *a posteriori* inferior, é necessária uma análise mais ampla. Os resultados poderiam estar associados aos exemplos numéricos que foram discutidos. No entanto, esse estudo preliminar comparativo apresentado não fornece evidências da capacidade de retenção de informação dos dados experimentais. Assim, é necessário um estudo para um melhor entendimento e comparação dos modelos. Na seção 4.7 será apresentada a descrição do processo de simulação de Monte Carlo que fornecerá subsídios para uma melhor comparação entre os modelos FVO e FVC. O motivo pelo qual somente o modelo de verossimilhança completa será usado na comparação com a metodologia proposta é que, na literatura, o modelo FVC é tido como padrão ouro, uma vez que utiliza toda a informação proveniente dos dados amostrais.

4.7. DELINEAMENTO DO PROCESSO DE SIMULAÇÃO

Para a execução do procedimento de simulação proposto, isto é, para a comparação da qualidade inferencial das distribuições *a posteriori* dos modelos FVO e FVC, foi elaborado um estudo que envolveu seis variáveis de entrada, a citar: o tamanho da amostra ou do experimento (n), o número de classificações repetidas de cada item (m), a proporção de itens conformes (p), os erros de classificação (e_1 e e_2) e o nível de informação da distribuição *a priori* de p . Desse modo, definiram-se os seguintes valores para esses parâmetros:

- a) $n = 50, 100$ e 150 ;
- b) $m = 3, 5$ e 7 ;
- c) $p = 0,55, 0,75$ e $0,95$;
- d) $(e_1; e_2) = (0,05 ; 0,05)$; $(e_1; e_2) = (0,15 ; 0,05)$ e $(e_1; e_2) = (0,15 ; 0,15)$
- e) Distribuição *a priori* para os erros de classificação: Beta(1,2) e Beta(2,10);

Para cada um desses níveis, a moda da distribuição *a posteriori* de p foi utilizada como preditor pontual dos modelos analisados. Além disso, os valores selecionados de n e de m servirão para dar uma idéia do comportamento das medidas inferenciais. Já os valores de p foram escolhidos como uma tentativa de representar, da forma mais realista possível, uma situação prática, em que a proporção de itens conformes refletiria três cenários produtivos: pessimista, moderado e otimista. Por outro lado, os dois níveis para os erros de classificação (5% e 15%) funcionarão como parâmetros perturbadores da precisão do sistema de classificação de cada item. E, por fim, as distribuições *a priori* desses erros auxiliará numa espécie de análise de sensibilidade em relação ao parâmetro p . Procurou-se, também, evitar uma distribuição muito informativa para os erros, adotando, apesar da constatação prática feita por Johnson *et al.* (1991) de que tais erros tendem a ser, em geral, menores que 50%.

A combinação desses seis fatores gerou 162 cenários possíveis para análise. E para evitar que a geração de um experimento amostral não favoreça a um ou ao outro modelo, em termos de desempenho preditivo, serão geradas, para cada um dos 162 cenários, ou seja, para cada vetor de entrada (n, m, p, e_1, e_2 , distribuição *a priori*), 500 repetições (k), totalizando a investigação de 81000 experimentos.

A partir desse arcabouço experimental, o método numérico de integração de Monte Carlo foi empregado para estimar a distribuição *a posteriori* de cada modelo para cada uma das k repetições e, em seguida, foram calculadas as medidas estatísticas de interesse para cada um dos modelos analisados, como destacadas abaixo:

1. Moda: Moda da distribuição *a posteriori* $f(p)$;
2. Desvio Padrão: Desvio padrão da distribuição *a posteriori* $f(p)$;

3. Amplitude do Intervalo MDP: Amplitude do intervalo de credibilidade de máxima densidade *a posteriori* (MDP) com probabilidade de 95%;
4. Razão dos Intervalos MDP: Razão das amplitudes dos intervalos MDP dos modelos FVO e FVC nessa ordem;
5. Fator de Bayes: Fator de Bayes do modelo FVO em relação ao modelo FVC;

Após a execução das $k=500$ repetições de cada cenário, foram obtidas a média aritmética de cada uma das medidas apresentadas acima. Nota-se, além do mais, que os referidos resultados, conjuntamente, auxiliam na comparação dos modelos em relação ao desempenho inferencial dos mesmos, sob a ótica das medidas de locação e de dispersão das distribuições *a posteriori*. Tais medidas ajudam a entender o quão próximo do verdadeiro valor do parâmetro p estão as estimativas pontuais bayesianas de cada modelo e com qual padrão de variabilidade. Sobre um âmbito mais geral, o algoritmo de simulação que descreve o procedimento descritivo de dados é mostrado na FIGURA 4.13.

ALGORITMO 2: COMPARAÇÃO DOS MODELOS FVO E FVC

Entrada: $n, m, p, e_1, e_2, \alpha, \text{Beta}(\lambda, \tau)$ e k

Saída: Moda, desvio padrão, amplitude do intervalo MDP e fator de Bayes

Para cada cenário $j=1$ até 162 faça:

- Para $i=1$ até k faça:
 - ▶ Gerar os dados a partir de $n, m, p, e_1, e_2, \text{Beta}(\alpha, \beta)$;
 - ▶ A partir do experimento, obter as quantidades $n_c, n_{\bar{c}}, C$ e \bar{C} ;
 - ▶ Obter as distribuições *a posteriori* de p dos modelos FVSO e MEB por meio do ALGORITMO 1 discutido na seção 4.2.2 e suas respectivas modas, desvios padrão, amplitudes dos intervalos de credibilidade e o fator de Bayes;
- Obter o valor médio de cada uma das quantidades calculadas no passo anterior;

FIGURA 4.13: Algoritmo de simulação para a comparação dos modelos FVO e FVC

Para comparar a quantidade de informação extraída dos dados experimentais por meio das distribuições preditivas de cada modelo, a metodologia utilizada terá como base o fator de Bayes. Como descrito na seção 4.5, o modelo M_1 representará a técnica proposta, enquanto que o modelo M_2 estará relacionado à técnica abordada por Evans *et al.* (1996). Além disso, a obtenção dessa quantidade obrigatoriamente passará pelo cálculo das distribuições preditivas de cada modelo, ou seja, pela aproximação das integrais $p(x|M_1)$ e $p(x|M_2)$. Isso significa que será necessário, mais uma vez, o uso dos métodos de integração de Monte Carlo. Em resumo, o cálculo do fator de Bayes, como foi proposto, dará uma ideia da qualidade do modelo M_1 em detrimento do modelo M_2 , em termos de absorção de informação, à luz dos dados observados. Obviamente outros critérios, tais como AIC e BIC, também poderiam ser utilizados.

Por fim, em termos de capacidade de máquina, foi utilizado, para a simulação de todos os resultados, o *Statistical Analysis System*[®] - SAS, versão 9.1.3, executando na plataforma Windows, sob uma configuração de um Pentium IV com processador Intel *duo-core* de 4 Gb de memória RAM. O esforço computacional, em termos de tempo, se apresentou de forma logarítmica em função do produto $n \times m$. Para se ter uma idéia, a estrutura mais simples com $n=50$ e $m=3$, os resultados foram gerados em, aproximadamente, 30 horas, enquanto que, para a estrutura com maior carga de trabalho ($n=150$ e $m=7$), aguardou-se cerca de oito dias para a obtenção de todas as medidas estatísticas comparativas. No total, o processo completo de simulação consumiu por volta de 37 dias. A FIGURA 4.14 mostra todos os procedimentos empregados na comparação dos modelos FVO e FVC para um cenário específico, mas considerando $k=500$ repetições.

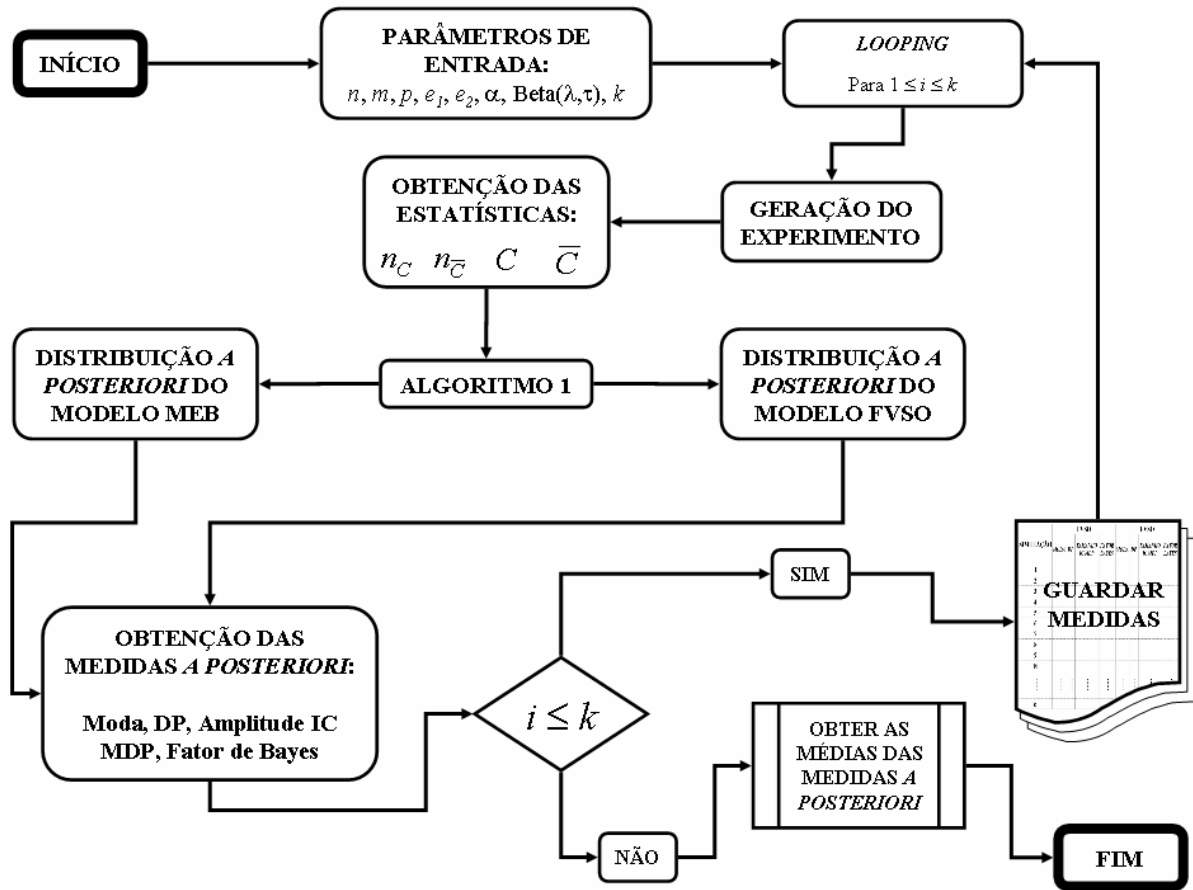


FIGURA 4.14: Tarefas realizadas para a comparação dos modelos FVO e FVC

4.8. RESULTADOS NUMÉRICOS

As TABELAS 4.6, 4.7 e 4.8 descrevem os resultados das simulações para os valores médios *a posteriori* modais e os desvios padrão para os experimentos com 50, 100 e 150 itens avaliados. Observa-se que os dois modelos se mostraram muito próximos quando se avaliou o vício. Os resultados sinalizam que o modelo FVO apresenta um menor vício quando foi usada a distribuição *a priori* Beta(1,2) para os erros de classificação. Enquanto o modelo FVC obteve um melhor desempenho com o uso da distribuição *a priori* Beta(2,10). Ao mesmo tempo, há indícios de que os

desvios padrão sofrem influências dos erros de classificação e do nível de informação *a priori* adotado. O desvio padrão *a posteriori* diminuiu com o aumento do tamanho da amostra e/ou do número de classificações repetidas. A FIGURA 4.15 ilustra a influência conjunta de m e n sobre a magnitude do desvio padrão dos modelos FVO e FVC, mostrando que tanto o número de classificações repetidas quanto o tamanho amostral do experimento contribuem, cada qual com o seu respectivo grau, diretamente para a diminuição gradativa dos desvios padrão de ambas as propostas de modelagem. Observa-se também que o modelo FVO apresenta um menor desvio quando comparado com o modelo FVC, principalmente quando a distribuição *a priori* Beta(1,2) é considerada. Isso mostra a possível importância da estrutura de ordenação na construção da função de verossimilhança proposta, discutida na seção 4.3. Finalmente, é necessário destacar que todos os casos simulados os desvios padrão *a posteriori* da FVO proposta foram inferiores ao da metodologia FVC, consistindo, assim, em um indicativo de superioridade.

Além disso, ainda pensando em dispersão dos modelos FVO e FVC, falta investigar os efeitos simultâneos do tamanho do experimento (n) com a proporção de itens conformes (p) e com os erros de classificação sobre os desvios padrão das duas propostas de modelagem inferencial. A FIGURA 4.16 mostra que não existem indícios significativos de que a proporção realmente influencie o desvio padrão dos modelos, muito embora se perceba que as respectivas distribuições *a posteriori* se apresentaram mais concentradas para taxas verdadeiras de conformidade maiores. Por outro lado, constata-se que os erros de classificação separados ou conjuntamente interferem na magnitude dos desvios padrão dos modelos analisados, provocando um aumento dessas medidas à proporção que os erros de julgamento crescem (FIGURA 4.17). A explicação para esse comportamento talvez esteja no fato de que a presença dos erros de classificação cause uma maior perturbação ou uma espécie de confusão no processo inferencial dos modelos, refletindo-se, conseqüentemente, em estimativas mais instáveis do parâmetro de interesse.

A conclusão mais importante dessa discussão até o momento é que os desvios padrão das estimativas do modelo proposto se mostraram sistematicamente menores do que os do modelo concorrente, quer seja quando comparados sob a ótica da proporção de itens conformes, quer seja quando avaliados sob o aspecto da influência dos erros de classificação.

TABELA 4.6: Valores modais e desvios padrão médios para os modelos proposto e FVC para $n=50$, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições *a priori*.

p	e_1	e_2	m	Nível de Informação da distribuição <i>a priori</i>							
				Beta(1,2)				Beta(2,10)			
				MODA		DESVIO PADRÃO		MODA		DESVIO PADRÃO	
				FVO	FVC	FVO	FVC	FVO	FVC	FVO	FVC
0,55	0,05	0,05	3	0,55033	0,54899	0,04802	0,05005	0,55114	0,54959	0,04704	0,04811
			5	0,55098	0,54940	0,04756	0,04900	0,54922	0,54826	0,04702	0,04828
			7	0,54991	0,54871	0,04736	0,04895	0,55074	0,55137	0,04692	0,04815
			3	0,54970	0,54831	0,04948	0,05187	0,54900	0,54905	0,04855	0,05040
			5	0,55076	0,54948	0,04782	0,04951	0,55045	0,54976	0,04670	0,04767
			7	0,55134	0,54947	0,04735	0,04869	0,54987	0,54868	0,04667	0,04756
	0,15	0,15	3	0,55563	0,54835	0,05394	0,05916	0,55615	0,54988	0,05044	0,05340
			5	0,55504	0,54793	0,04929	0,05162	0,55590	0,54905	0,04738	0,04857
			7	0,55413	0,54839	0,04842	0,05030	0,55500	0,55151	0,04658	0,04741
			3	0,74786	0,74742	0,04377	0,04647	0,74919	0,74932	0,04152	0,04297
			5	0,74943	0,74899	0,04285	0,04560	0,74871	0,74823	0,04125	0,04266
			7	0,74872	0,74971	0,04245	0,04464	0,74969	0,74995	0,04116	0,04254
0,75	0,15	0,05	3	0,74689	0,74711	0,04814	0,05295	0,74709	0,74781	0,04380	0,04596
			5	0,74815	0,74770	0,04519	0,04847	0,74900	0,74912	0,04105	0,04204
			7	0,74882	0,74899	0,04467	0,04778	0,74929	0,74931	0,04102	0,04213
			3	0,74919	0,74187	0,05116	0,05826	0,75003	0,74577	0,04478	0,04762
			5	0,75141	0,74623	0,04822	0,05419	0,75261	0,74886	0,04186	0,04351
			7	0,75099	0,74689	0,04790	0,05350	0,75260	0,74939	0,04090	0,04201
	0,05	0,05	3	0,94544	0,93777	0,02886	0,03550	0,94799	0,94833	0,02157	0,02222
			5	0,94568	0,93831	0,02710	0,03262	0,94840	0,94915	0,02061	0,02064
			7	0,94451	0,94079	0,02668	0,03137	0,94866	0,94940	0,02034	0,02109
			3	0,94252	0,93370	0,03127	0,03977	0,94458	0,94161	0,02520	0,02582
			5	0,94353	0,92258	0,03120	0,03784	0,94739	0,94677	0,02253	0,02315
			7	0,94159	0,93481	0,03044	0,03827	0,94791	0,94839	0,02150	0,02191
0,95	0,15	0,15	3	0,94396	0,92692	0,03272	0,04345	0,94481	0,93805	0,02540	0,02618
			5	0,94312	0,91347	0,03220	0,04054	0,94831	0,94602	0,02224	0,02356
			7	0,94212	0,92566	0,03141	0,04185	0,94879	0,94743	0,02128	0,02222

TABELA 4.7: Valores modais e desvios padrão médios para os modelos proposto e FVC para $n=100$, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições *a priori*.

p	e_1	e_2	m	Nível de Informação da distribuição <i>a priori</i>							
				Beta(1,2)				Beta(2,10)			
				MODA		DESVIO PADRÃO		MODA		DESVIO PADRÃO	
				FVO	FVC	FVO	FVC	FVO	FVC	FVO	FVC
0,55	0,05	0,05	3	0,55010	0,54906	0,03518	0,03755	0,54998	0,54943	0,03430	0,03557
			5	0,55139	0,54932	0,03491	0,03648	0,55037	0,54934	0,03414	0,03530
			7	0,55066	0,54846	0,03461	0,03648	0,55078	0,54949	0,03407	0,03501
	0,15	0,05	3	0,54899	0,54882	0,03623	0,03832	0,54983	0,54995	0,03510	0,03649
			5	0,55108	0,55004	0,03517	0,03706	0,55110	0,54972	0,03423	0,03536
			7	0,55048	0,54882	0,03495	0,03660	0,55083	0,55008	0,03393	0,03483
	0,15	0,15	3	0,55578	0,54951	0,04017	0,04483	0,55596	0,54979	0,03769	0,04068
			5	0,55617	0,54769	0,03603	0,03803	0,55645	0,55042	0,03430	0,03524
			7	0,55431	0,54931	0,03580	0,03776	0,55449	0,55039	0,03377	0,03445
0,75	0,05	0,05	3	0,74846	0,74790	0,03186	0,03410	0,74893	0,74903	0,03011	0,03138
			5	0,75049	0,74978	0,03153	0,03401	0,74949	0,74947	0,03009	0,03114
			7	0,74865	0,74904	0,03126	0,03355	0,74939	0,74968	0,02957	0,03053
	0,15	0,05	3	0,74672	0,74776	0,03581	0,04032	0,74651	0,74786	0,03188	0,03379
			5	0,74807	0,74885	0,03412	0,03784	0,74910	0,74986	0,02990	0,03081
			7	0,74891	0,74816	0,03356	0,03709	0,74950	0,74975	0,02972	0,03068
	0,15	0,15	3	0,74971	0,74466	0,04169	0,05041	0,75093	0,74847	0,03351	0,03651
			5	0,75133	0,74387	0,03874	0,04583	0,75233	0,74935	0,03040	0,03152
			7	0,75167	0,74387	0,03781	0,04432	0,75293	0,74983	0,02960	0,03038
0,95	0,05	0,05	3	0,94674	0,94534	0,02199	0,02683	0,94776	0,94823	0,01597	0,01666
			5	0,94701	0,94755	0,01886	0,02189	0,94847	0,94883	0,01517	0,01582
			7	0,94640	0,94791	0,01733	0,01938	0,94890	0,94945	0,01489	0,01552
	0,15	0,05	3	0,94219	0,93330	0,02654	0,03447	0,94434	0,94387	0,01844	0,01941
			5	0,94519	0,94466	0,02348	0,02957	0,94752	0,94825	0,01611	0,01648
			7	0,94476	0,94520	0,02211	0,02709	0,94838	0,94968	0,01540	0,01562
	0,15	0,15	3	0,94482	0,93739	0,02623	0,03634	0,94516	0,94311	0,01814	0,01933
			5	0,94398	0,92330	0,02585	0,03473	0,94821	0,94756	0,01634	0,01706
			7	0,94394	0,93907	0,02560	0,03551	0,94929	0,94916	0,01520	0,01580

TABELA 4.8: Valores modais e desvios padrão médios para os modelos proposto e FVC para $n=150$, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições *a priori*.

p	e_1	e_2	m	Nível de Informação da distribuição <i>a priori</i>							
				Beta(1,2)				Beta(2,10)			
				MODA		DESVIO PADRÃO		MODA		DESVIO PADRÃO	
				FVO	FVC	FVO	FVC	FVO	FVC	FVO	FVC
0,55	0,05	0,05	3	0,54987	0,54844	0,02896	0,03117	0,55094	0,54967	0,02841	0,02963
			5	0,55075	0,54923	0,02881	0,03030	0,54958	0,54961	0,02824	0,02955
			7	0,55063	0,55034	0,02865	0,03049	0,55085	0,54989	0,02817	0,02912
	0,15	0,05	3	0,54962	0,55105	0,02988	0,03167	0,54922	0,55061	0,02903	0,03029
			5	0,55053	0,54968	0,02932	0,03153	0,55104	0,55029	0,02826	0,02936
			7	0,55052	0,54896	0,02924	0,03123	0,55025	0,54930	0,02821	0,02917
	0,15	0,15	3	0,55548	0,54879	0,03395	0,03846	0,55579	0,54895	0,03154	0,03433
			5	0,55673	0,54889	0,03042	0,03268	0,55670	0,55020	0,02834	0,02920
			7	0,55426	0,54871	0,02988	0,03189	0,55486	0,55019	0,02803	0,02879
0,75	0,05	0,05	3	0,74894	0,74972	0,02620	0,02888	0,74923	0,74982	0,02510	0,02652
			5	0,74940	0,74956	0,02618	0,02821	0,74950	0,74955	0,02486	0,02586
			7	0,74876	0,74946	0,02574	0,02798	0,75051	0,75024	0,02448	0,02550
	0,15	0,05	3	0,74521	0,74617	0,03045	0,03515	0,74689	0,74865	0,02635	0,02804
			5	0,74937	0,75000	0,02802	0,03136	0,74921	0,74919	0,02468	0,02556
			7	0,74846	0,74884	0,02744	0,03056	0,74973	0,74983	0,02456	0,02547
	0,15	0,15	3	0,74944	0,74425	0,03638	0,04585	0,75073	0,74886	0,02827	0,03121
			5	0,75184	0,74491	0,03329	0,04052	0,75245	0,74909	0,02497	0,02589
			7	0,75169	0,74323	0,03280	0,03983	0,75276	0,74914	0,02452	0,02528
0,95	0,05	0,05	3	0,94661	0,94818	0,01761	0,02130	0,94822	0,94883	0,01310	0,01386
			5	0,94724	0,94758	0,01488	0,01767	0,94885	0,94938	0,01239	0,01305
			7	0,94583	0,94760	0,01423	0,01657	0,94917	0,94975	0,01227	0,01296
	0,15	0,05	3	0,94307	0,93816	0,02512	0,03475	0,94479	0,94534	0,01562	0,01694
			5	0,94520	0,94531	0,01998	0,02504	0,94751	0,94889	0,01334	0,01380
			7	0,94565	0,94463	0,01831	0,02324	0,94878	0,94967	0,01256	0,01299
	0,15	0,15	3	0,94402	0,92876	0,02544	0,03778	0,94538	0,94472	0,01590	0,01742
			5	0,94467	0,94013	0,02403	0,03440	0,94855	0,94857	0,01351	0,01427
			7	0,94569	0,94182	0,02197	0,03129	0,94943	0,94931	0,01264	0,01331

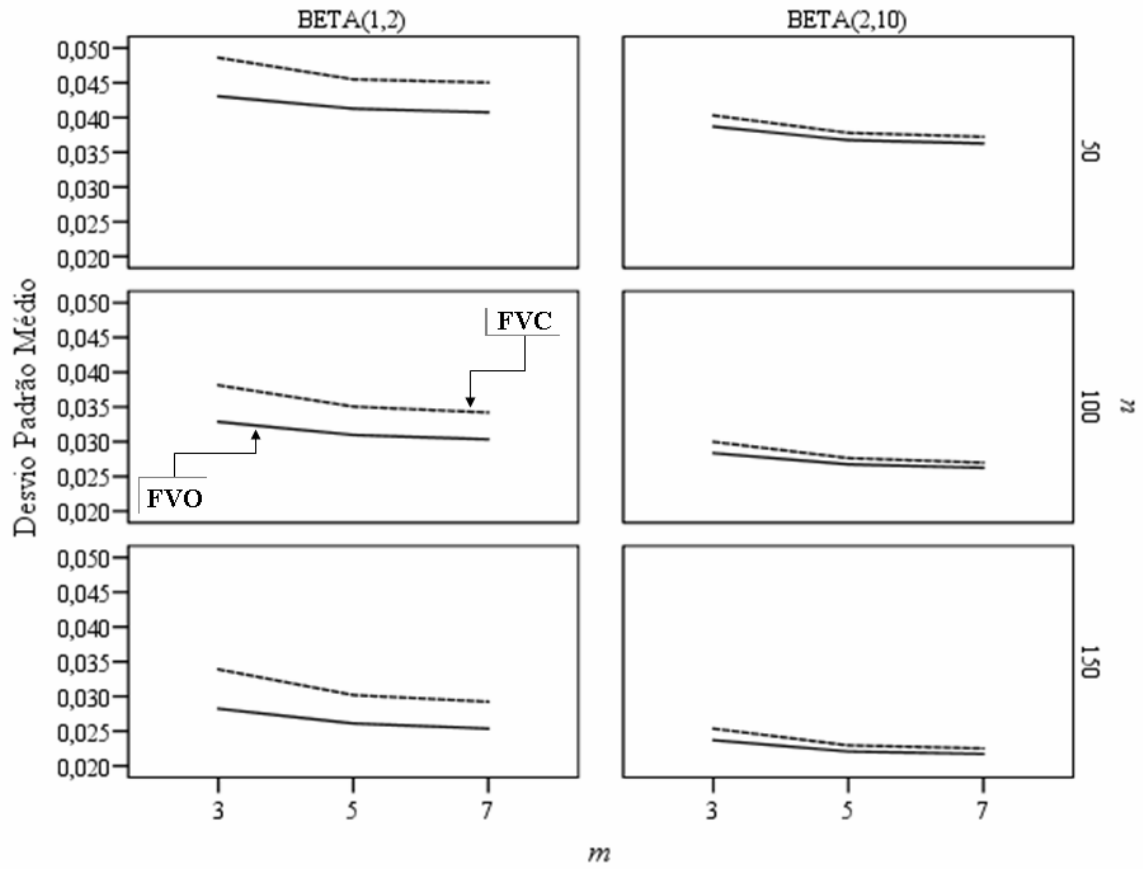


FIGURA 4.15: Comparação dos modelos FVO e FVC por meio do desvio padrão médio, segundo a quantidade de informação *a priori*, mas controlando pelo tamanho da amostra e pelo número de classificações repetidas

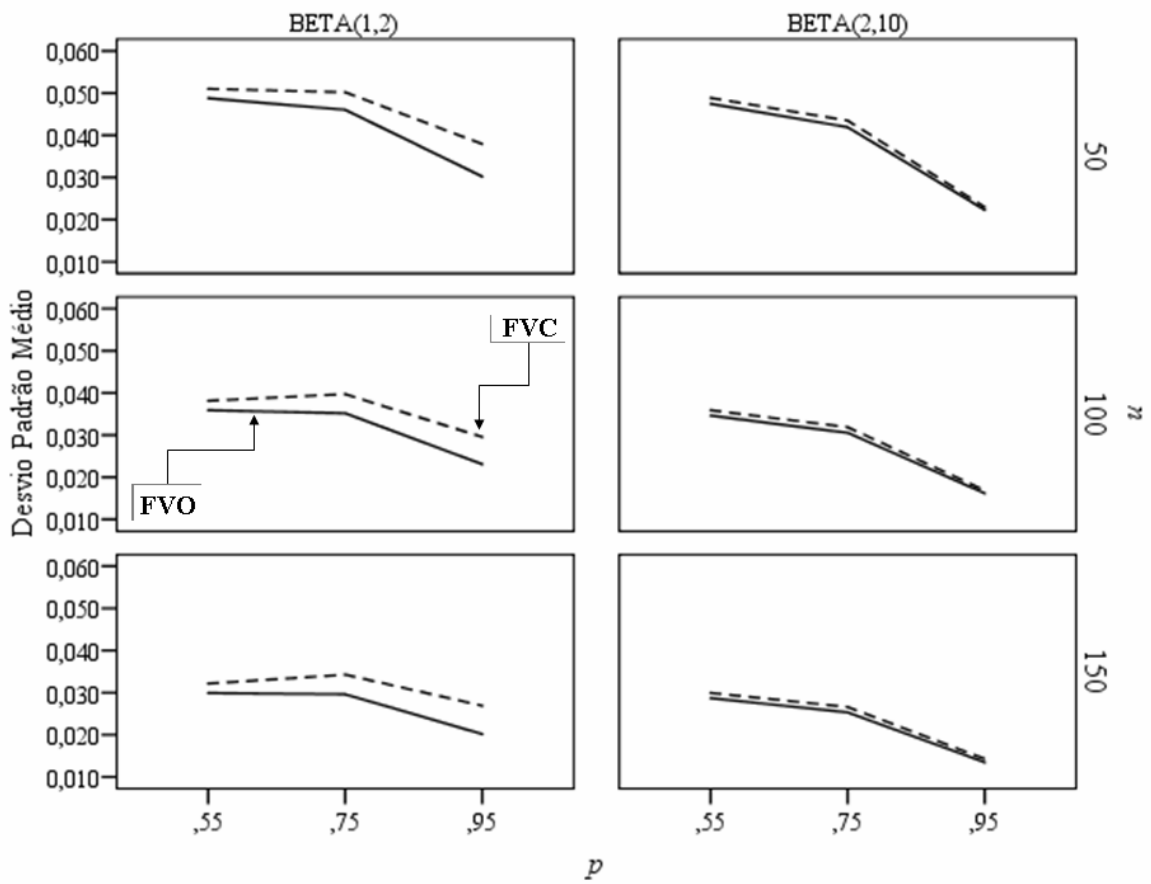


FIGURA 4.16: Comparação dos modelos FVO e FVC por meio do desvio padrão, segundo a quantidade de informação α *priori*, mas controlando pelo tamanho da amostra e pela proporção de itens conformes

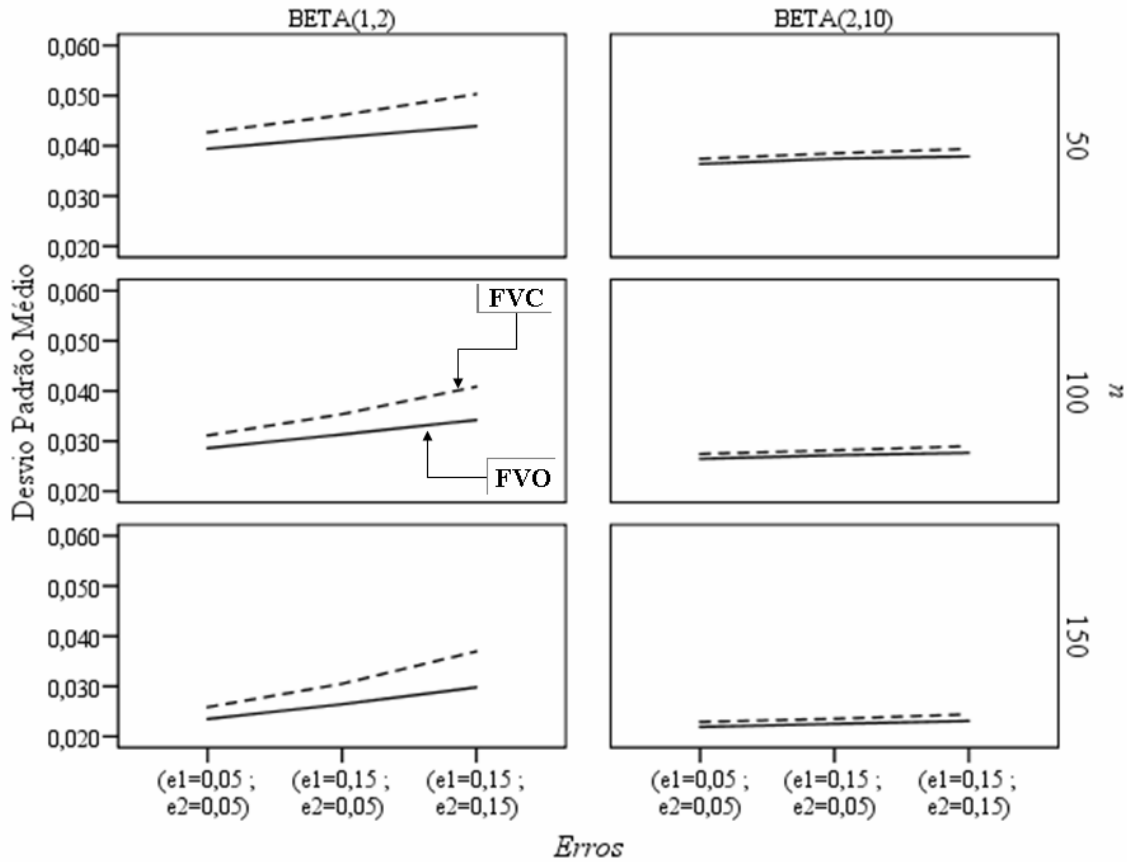


FIGURA 4.17: Comparação dos modelos FVO e FVC por meio do desvio padrão, segundo a quantidade de informação *a priori*, mas controlando pelo tamanho da amostra e pelos erros de classificação

Conforme foi previsto no delineamento do processo de simulação, além da comparação dos modelos FVO e FVC por meio suas medidas descritivas *a posteriori* de centralidade e de variabilidade, planejou-se também um procedimento comparativo inferencial desses modelos com base na razão das amplitudes dos respectivos intervalos de máxima densidade *a posteriori* com 95% de credibilidade e do cálculo do fator de Bayes. Essas duas novas medidas juntas servirão para gerar novos subsídios na escolha entre os modelos FVO e FVC.

Desse modo, as TABELAS 4.9, 4.10 e 4.11 relatam os valores médios das amplitudes dos intervalos de 95% de credibilidade e do fator de Bayes para experimentos amostrais com 50, 100 e

150 itens. Numa leitura vertical sobre cada tabela, observa-se que as amplitudes dos intervalos de ambos os modelos, para cada grau de informação da distribuição *a priori* dos erros e para de cada faixa da proporção de itens conformes, aumentam à medida que são considerados erros de julgamentos maiores, quer seja do erro tipo I, quer seja do erro tipo II. Já se baseando num recorte por linha, nota-se que o uso de distribuições *a priori* mais informativas para os erros de classificação contribuem para projetar distribuições *a posteriori* com intervalos de credibilidade mais curtos para os dois modelos, ou seja, mais concentradas.

Além disso, percebe-se que tanto o tamanho da amostra quanto o número de classificações repetidas considerados ajudam na redução dos desvios padrão dos estimadores dos modelos. Mas, independente do ponto de vista que for utilizado, comprova-se que as amplitudes dos intervalos do modelo FVO são sempre menores que os tamanhos dos intervalos do modelo investigado por Evans *et al.* (1996). Essa constatação conduz a mais um resultado importante, pois mais uma vez sinaliza a superioridade do procedimento proposto. Tal fato pode ser comprovado na razão dos intervalos observados que são inferiores a 1, para quaisquer níveis de n , m , p , e_1 , e_2 e quantidades de informação das distribuições *a priori* (TABELAS 4.9, 4.10 e 4.11).

Sob um ponto de vista mais estocástico, é possível analisar as características dessas razões em termos de suas medidas de centralidade, variabilidade, assimetria e pontos discrepantes. Por exemplo, a FIGURA 4.18 mostra as distribuições amostrais das razões das amplitudes dos intervalos de credibilidade controlando por n , m e o nível de informação da distribuição *a priori* dos erros. Nesse gráfico é possível perceber que o uso das distribuições *a priori* Beta(1,2) para os erros geram distribuições assimétricas e dispersas para as razões das amplitudes dos intervalos de credibilidade, enquanto que as distribuições *a priori* Beta(2,10) produzem distribuições ligeiramente simétricas, bem mais concentradas e mais próximas de 1. Essa última afirmação retrata uma clara tendência dos modelos apresentarem amplitudes dos intervalos de credibilidade semelhantes na presença de

distribuições *a priori* Beta(2,10), ou seja, quanto mais informação *a priori* for considerada, mais próximos serão os perfis inferenciais dos modelos. Ainda assim, como pode ser comprovado na própria FIGURA 4.19, mesmo havendo uma proximidade das amplitudes dos intervalos de credibilidade dos modelos bayesianos FVO e FVC, há de se destacar que as amplitudes dos intervalos de credibilidade geradas pela técnica proposta foram sempre menores que as amplitudes dos intervalos produzidos pela última técnica, independente da situação trabalhada.

Adicionalmente, observa-se que o aumento do tamanho da amostra e/ou do número de classificações repetidas não interferiu na relação proporcional dos tamanhos dos intervalos de credibilidade dos modelos. A mesma conclusão pode ser aplicada àqueles gráficos que envolveram a proporção de itens conformes (FIGURA 4.19) e os erros de classificação (FIGURA 4.20), ou seja, apesar de as distribuições das razões dos intervalos se apresentarem com diferentes perfis de assimetria e de variabilidade, mostrando uma ausência de uma relação direta, o fator com maior poder de influência sobre tais razões foi o grau de informação das distribuições *a priori*.

TABELA 4.9: Amplitudes médias dos intervalos de 95% de credibilidade para os modelos propostos e FVC e fator de Bayes para $n=50$, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições *a priori*.

p	e_1	e_2	m	Nível de Informação da Distribuição <i>a priori</i>							
				Beta(1,2)			Beta(2,10)				
				AMPLITUDE DO INTERVALO		RAZÃO (A/B)	FATOR DE BAYES	AMPLITUDE DO INTERVALO		RAZÃO (A/B)	FATOR DE BAYES
				FVO (A)	FVC (B)			FVO (A)	FVC (B)		
0,55	0,05	0,05	3	0,11292	0,11766	0,95967	22,75117	0,11061	0,11316	0,97748	2,34019
			5	0,11179	0,11521	0,97035	22,21845	0,11056	0,11354	0,97371	2,16517
			7	0,11136	0,11512	0,96737	22,81505	0,11035	0,11325	0,97439	2,29578
	0,15	0,05	3	0,11635	0,12193	0,95430	22,58916	0,11412	0,11854	0,96277	2,16647
			5	0,11241	0,11643	0,96543	21,48893	0,10984	0,11212	0,97968	2,43115
			7	0,11135	0,11436	0,97371	21,87516	0,10976	0,11185	0,98136	2,16741
	0,15	0,15	3	0,12669	0,13879	0,91282	21,99032	0,11850	0,12541	0,94490	2,27915
			5	0,11585	0,12129	0,95511	22,76263	0,11142	0,11424	0,97530	2,28587
			7	0,11383	0,11821	0,96296	21,07648	0,10955	0,11150	0,98248	2,24439
0,75	0,05	0,05	3	0,10182	0,10811	0,94179	23,39429	0,09683	0,10045	0,96392	2,23189
			5	0,09937	0,10576	0,93952	23,04235	0,09644	0,09972	0,96711	2,30246
			7	0,09888	0,10411	0,94978	22,10038	0,09620	0,09954	0,96647	2,31793
	0,15	0,05	3	0,11167	0,12100	0,92286	23,28123	0,10254	0,10710	0,95745	2,29841
			5	0,10506	0,11207	0,93748	23,05175	0,09615	0,09847	0,97648	2,20294
			7	0,10395	0,11069	0,93909	22,45494	0,09598	0,09862	0,97315	2,13830
	0,15	0,15	3	0,11740	0,13045	0,89998	21,21010	0,10452	0,11041	0,94668	2,23451
			5	0,11040	0,12304	0,89728	23,10625	0,09757	0,10174	0,95895	2,26309
			7	0,10983	0,12170	0,90245	22,35149	0,09552	0,09840	0,97075	2,25349
0,95	0,05	0,05	3	0,05812	0,06816	0,85269	22,33988	0,04518	0,04519	0,99958	2,21823
			5	0,05456	0,06273	0,86967	22,88025	0,04201	0,04248	0,98904	2,29020
			7	0,05435	0,06051	0,89821	22,01462	0,04217	0,04243	0,99373	2,27288
	0,15	0,05	3	0,06217	0,07419	0,83797	22,61712	0,05302	0,05539	0,95723	2,26955
			5	0,06526	0,07439	0,87728	23,32456	0,04733	0,04791	0,98793	2,27154
			7	0,06088	0,07200	0,84558	22,18172	0,04470	0,04539	0,98489	2,25518
	0,15	0,15	3	0,06366	0,08069	0,78894	22,17444	0,05400	0,05535	0,97573	2,23496
			5	0,06708	0,08035	0,83481	21,12922	0,04625	0,04814	0,96089	2,34942
			7	0,06098	0,07807	0,78119	21,83387	0,04412	0,04579	0,96345	2,24048

TABELA 4.10: Amplitudes médias dos intervalos de 95% de credibilidade para os modelos propostos e FVC e fator de Bayes para $n=100$, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições *a priori*.

p	e_1	e_2	m	Nível de Informação da Distribuição <i>a priori</i>							
				Beta(1,2)			Beta(2,10)				
				AMPLITUDE DO INTERVALO		RAZÃO (A/B)	FATOR DE BAYES	AMPLITUDE DO INTERVALO		RAZÃO (A/B)	FATOR DE BAYES
				FVO (A)	FVC (B)			FVO (A)	FVC (B)		
0,55	0,05	0,05	3	0,08273	0,08830	0,93699	22,56581	0,08070	0,08365	0,96467	2,38643
			5	0,08209	0,08579	0,95686	23,44966	0,08027	0,08301	0,96689	2,25236
			7	0,08140	0,08580	0,94879	22,50841	0,08012	0,08234	0,97300	2,24622
	0,15	0,05	3	0,08523	0,09011	0,94577	22,02143	0,08254	0,08582	0,96179	2,23685
			5	0,08265	0,08716	0,94827	22,11444	0,08047	0,08316	0,96761	2,19342
			7	0,08221	0,08609	0,95496	22,33047	0,07982	0,08193	0,97425	2,34069
	0,15	0,15	3	0,09446	0,10538	0,89636	20,58269	0,08868	0,09568	0,92684	2,25674
			5	0,08475	0,08944	0,94753	23,47810	0,08069	0,08288	0,97355	2,24760
			7	0,08421	0,08881	0,94819	22,67493	0,07945	0,08103	0,98042	2,31004
0,75	0,05	0,05	3	0,07476	0,07997	0,93488	22,04912	0,07076	0,07377	0,95922	2,15298
			5	0,07383	0,07933	0,93069	21,84334	0,07070	0,07318	0,96608	2,31067
			7	0,07335	0,07859	0,93338	23,02222	0,06947	0,07177	0,96795	2,29008
	0,15	0,05	3	0,08402	0,09405	0,89339	22,46335	0,07497	0,07941	0,94402	2,22075
			5	0,07989	0,08789	0,90907	21,64325	0,07033	0,07244	0,97087	2,34828
			7	0,07862	0,08621	0,91194	21,46844	0,06984	0,07213	0,96824	2,34290
	0,15	0,15	3	0,09712	0,11507	0,84398	21,22749	0,07871	0,08570	0,91839	2,28507
			5	0,09003	0,10493	0,85803	22,25175	0,07144	0,07410	0,96413	2,29166
			7	0,08797	0,10174	0,86467	23,45263	0,06952	0,07143	0,97319	2,17213
0,95	0,05	0,05	3	0,04632	0,05150	0,89937	21,18353	0,03574	0,03661	0,97619	2,31159
			5	0,04000	0,04333	0,92312	22,49390	0,03370	0,03484	0,96730	2,26817
			7	0,03741	0,03978	0,94055	22,15501	0,03287	0,03395	0,96815	2,20127
	0,15	0,05	3	0,05718	0,06539	0,87447	24,42144	0,04229	0,04241	0,99705	2,22234
			5	0,04917	0,05567	0,88334	22,80393	0,03619	0,03647	0,99230	2,37567
			7	0,04627	0,05114	0,90492	22,06936	0,03400	0,03455	0,98401	2,24000
	0,15	0,15	3	0,05283	0,06547	0,80695	22,50083	0,04077	0,04216	0,96716	2,20622
			5	0,05502	0,06683	0,82331	23,57212	0,03650	0,03725	0,97992	2,27478
			7	0,05076	0,06362	0,79783	21,79319	0,03341	0,03447	0,96916	2,22626

TABELA 4.11: Amplitudes médias dos intervalos de 95% de credibilidade para os modelos propostos e FVC e fator de Bayes para $n=150$, segundo a proporção de itens conformes, erros de classificação, número de classificações repetidas e níveis de informação das distribuições *a priori*.

p	e_1	e_2	m	Nível de Informação da Distribuição <i>a priori</i>							
				Beta(1,2)			Beta(2,10)				
				AMPLITUDE DO INTERVALO		RAZÃO (A/B)	FATOR DE BAYES	AMPLITUDE DO INTERVALO		RAZÃO (A/B)	FATOR DE BAYES
				FVO (A)	FVC (B)			FVO (A)	FVC (B)		
0,55	0,05	0,05	3	0,06809	0,07330	0,92894	22,42637	0,06682	0,06969	0,95877	2,21133
			5	0,06774	0,07126	0,95062	23,05762	0,06643	0,06949	0,95598	2,17532
			7	0,06740	0,07172	0,93975	22,76802	0,06623	0,06849	0,96694	2,26740
	0,15	0,05	3	0,07026	0,07450	0,94313	22,69878	0,06828	0,07125	0,95841	2,38135
			5	0,06893	0,07414	0,92972	23,24085	0,06646	0,06904	0,96264	2,29776
			7	0,06873	0,07345	0,93572	22,97089	0,06633	0,06861	0,96685	2,23667
	0,15	0,15	3	0,07984	0,09045	0,88274	23,06771	0,07417	0,08074	0,91866	2,22697
			5	0,07154	0,07686	0,93069	20,80132	0,06664	0,06868	0,97028	2,27901
			7	0,07023	0,07499	0,93656	22,04285	0,06595	0,06771	0,97395	2,40955
0,75	0,05	0,05	3	0,06141	0,06752	0,90945	22,32462	0,05901	0,06235	0,94650	2,30257
			5	0,06152	0,06626	0,92857	21,04551	0,05848	0,06081	0,96158	2,30442
			7	0,06050	0,06571	0,92075	22,14011	0,05761	0,05998	0,96060	2,28753
	0,15	0,05	3	0,07157	0,08204	0,87242	22,08658	0,06194	0,06596	0,93912	2,36060
			5	0,06578	0,07330	0,89741	22,54287	0,05802	0,06011	0,96517	2,38233
			7	0,06437	0,07121	0,90401	21,75887	0,05777	0,05990	0,96444	2,14693
	0,15	0,15	3	0,08502	0,10538	0,80683	22,94053	0,06648	0,07337	0,90611	2,25707
			5	0,07772	0,09367	0,82970	22,67619	0,05868	0,06089	0,96376	2,29120
			7	0,07647	0,09085	0,84168	23,37135	0,05763	0,05946	0,96922	2,19944
0,95	0,05	0,05	3	0,03825	0,04278	0,89420	23,36369	0,02981	0,03140	0,94923	2,24248
			5	0,03251	0,03679	0,88363	22,50213	0,02811	0,02946	0,95442	2,30762
			7	0,03130	0,03478	0,90013	22,34497	0,02765	0,02921	0,94648	2,26394
	0,15	0,05	3	0,05407	0,06438	0,83978	22,66093	0,03635	0,03757	0,96761	2,40990
			5	0,04325	0,04819	0,89755	22,43573	0,03068	0,03102	0,98891	2,19930
			7	0,03897	0,04515	0,86317	20,67704	0,02864	0,02932	0,97677	2,36356
	0,15	0,15	3	0,05353	0,06994	0,76541	22,07867	0,03670	0,03786	0,96925	2,30585
			5	0,04921	0,06174	0,79697	23,16125	0,03072	0,03198	0,96059	2,35891
			7	0,04430	0,05667	0,78164	21,63617	0,02851	0,03003	0,94926	2,20058

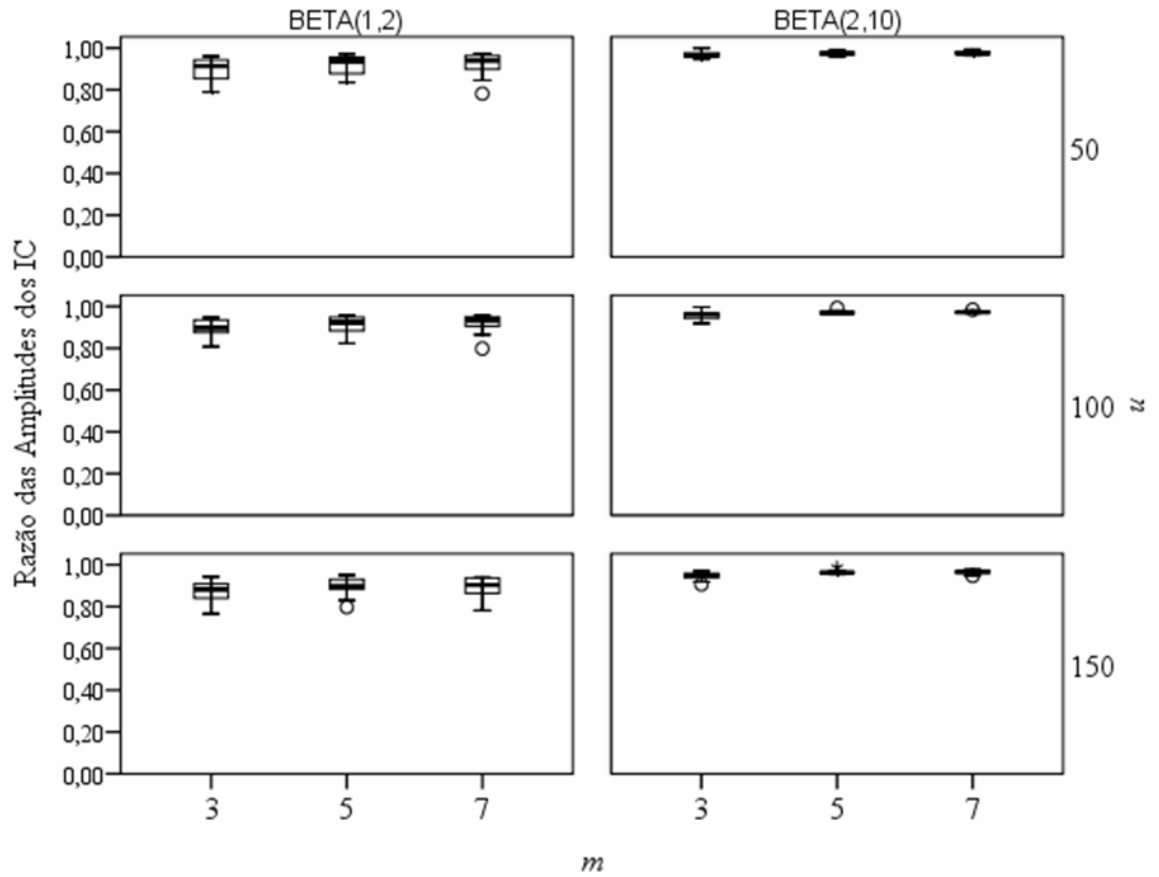


FIGURA 4.18: Distribuição da razão das amplitudes dos intervalos de 95% de credibilidade do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação *a priori*, mas controlando pelo tamanho da amostra e pelo número de classificações repetidas

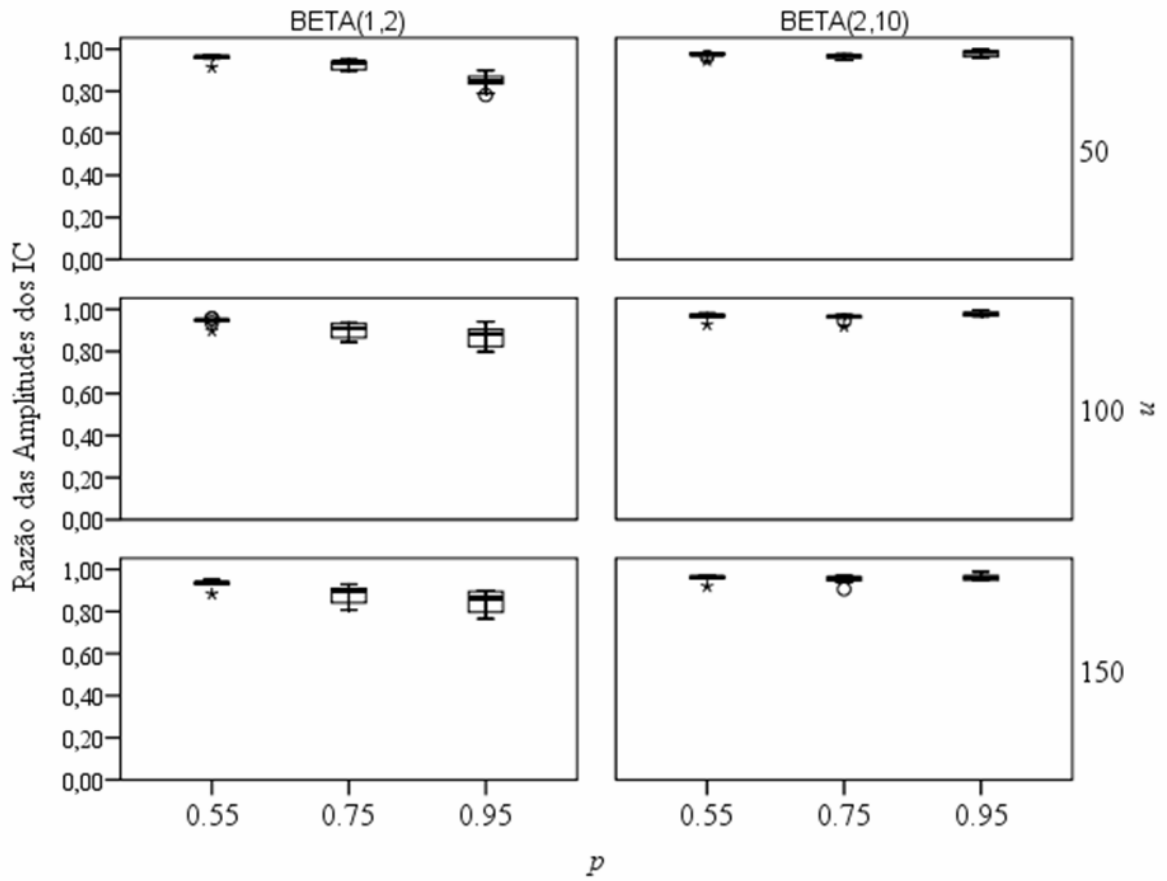


FIGURA 4.19: Distribuição da razão das amplitudes dos intervalos de 95% de credibilidade do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação *a priori*, mas controlando pelo tamanho da amostra e pela proporção de itens conformes

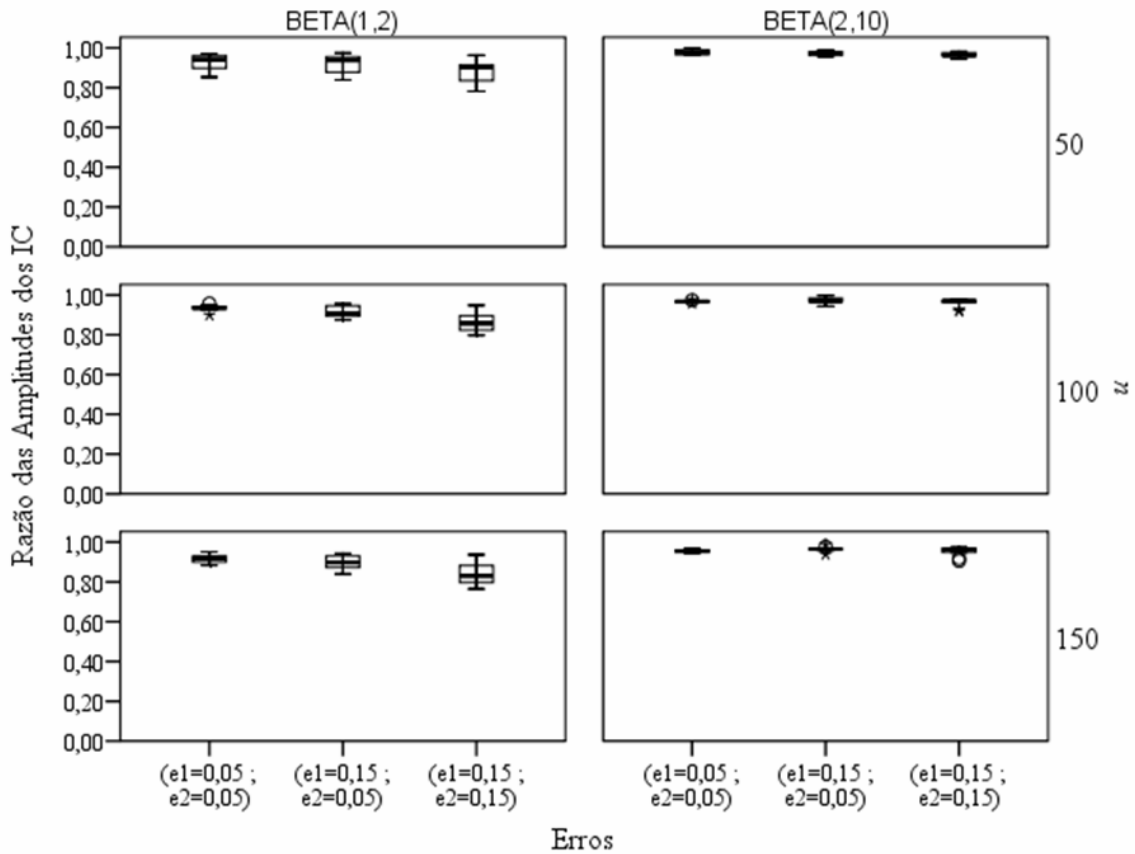


FIGURA 4.20: Distribuição da razão das amplitudes dos intervalos de 95% de credibilidade do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação *a priori*, mas controlando pelo tamanho da amostra e pelos os erros de classificação

Concomitantemente aos comentários ditos anteriormente, os modelos FVO e FVC foram comparados segundo a quantidade de informação amostral absorvida dos dados e medida por meio do fator de Bayes (ver seção 4.5). Para isso, produziu-se a mesma estrutura de simulação utilizada para investigar a razão dos intervalos de credibilidade dos modelos, mas rapidamente se percebeu, com exceção do nível de informação da distribuição *a priori* dos erros de julgamento, que nenhum dos fatores n , m , p , e_1 e e_2 apresentou um grau de explicação razoável do fator de Bayes. Por isso, resolveu-se construir apenas uma estrutura gráfica envolvendo somente o grau de informação da

distribuição *a priori*, n e m , sendo esses últimos os parâmetros responsáveis, geralmente, pelo controle das características assintóticas das distribuições *a posteriori*.

Assim, com base na FIGURA 4.21, nota-se que as distribuições amostrais dos fatores de Bayes se apresentaram bastante concentradas e oscilando em torno de 22,38, para distribuições *a priori* pouco informativas, e por volta de 2,27 para distribuições *a priori* com um grau moderado de informação sobre os parâmetros e_1 e e_2 . Observe, ainda, que essa afirmação é independente da influência dos valores de n e m . Mais que isso, essa análise demonstra que o modelo FVO, para uma distribuição *a priori* Beta(1,2), conseguiu apresentar fator de Bayes 22 vezes maior do que o modelo FVC. Isso, sobretudo, refletiu a capacidade que o modelo de verossimilhança ordenada tem de transformar uma informação amostral numa informação inferencial. De acordo com Jeffreys (1961) um fator de Bayes entre 10 e 100 é considerado uma forte evidência e entre 1 e 3,2 é considerado evidência muito fraca. Assim, com o uso da distribuição *a priori* Beta(1,2) fica evidenciada a superioridade do modelo proposto. Já, com uso da distribuição *a priori* Beta (2,10) a evidência de superioridade do modelo proposto não é tão evidente. Mas deve-se considerar o fato de que em todos os casos simulados o fator de Bayes foi favorável ao modelo proposto quando utilizou-se a distribuição *a priori* Beta (2,10).

No conjunto, considerando os desvios padrão, as razões das amplitudes dos intervalos de credibilidade e os fatores de Bayes, houve fortes indícios de que o modelo defendido aqui é superior ao modelo tratado em Evans *et al.* (1996).

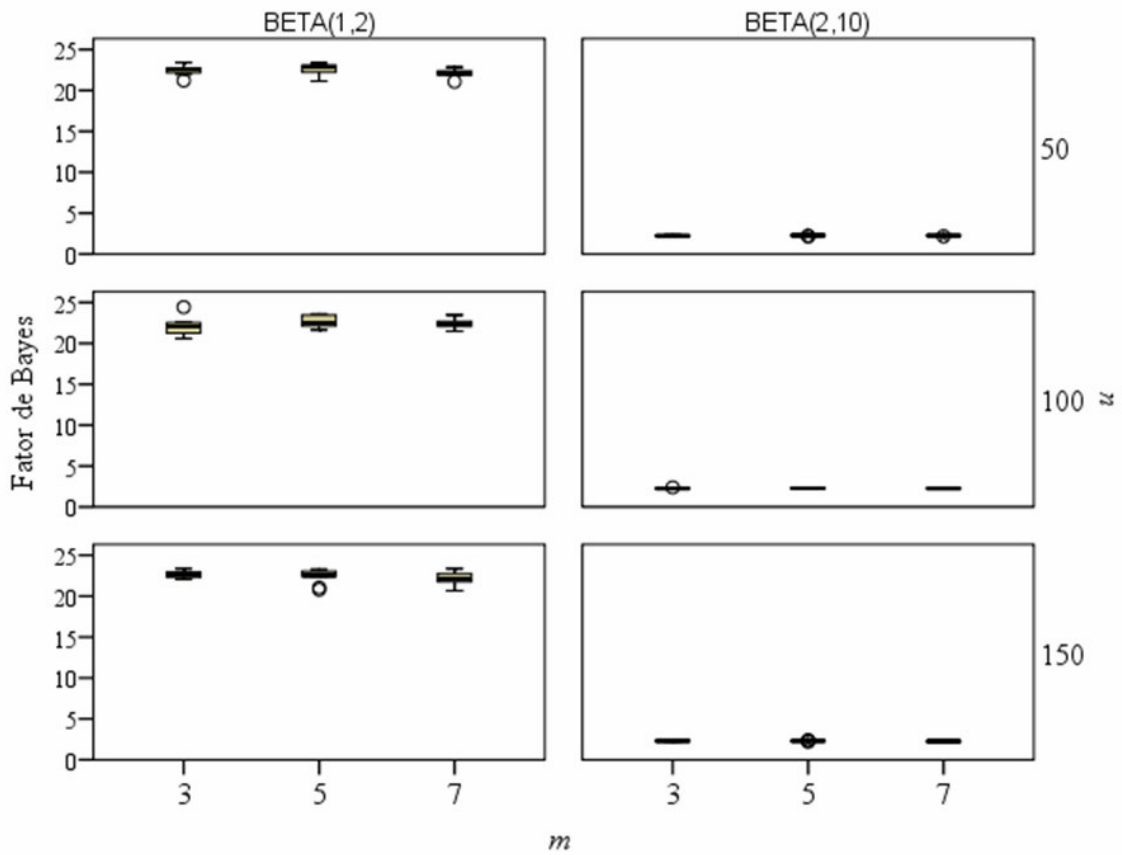


FIGURA 4.21: Distribuição do fator de Bayes do modelo FVO em relação ao modelo FVC, segundo a quantidade de informação *a priori*, mas controlado pelo tamanho da amostra e pelo número de classificações repetidas

No capítulo que segue outro importante tema será abordado: a obtenção dos valores ótimos de n , m e do critério para a classificação final do item (a). Até então tal critério de classificação final considerado nos modelos investigados foi $a=m/2$. Agora a proposta é relaxar essa condição para $1 \leq a \leq m$ e, conjuntamente com as variáveis de controle n e m , encontrar o vetor ótimo (n, m, a) que atenda a determinadas especificações inferenciais *a posteriori*. A ferramenta de otimização sugerida para realizar a tarefa descrita será um algoritmo genético híbrido multiobjetivo.

Capítulo 5

OTIMIZAÇÃO MULTIOBJETIVO VIA METAHEURÍSTICAS HÍBRIDAS

Este capítulo consiste em obter o tamanho da amostra (n), o número de classificações repetidas (m) e o critério (a) para a classificação final de cada item após as m classificações repetidas, de modo que esses, conjuntamente, atendam a especificações inferenciais de controle referentes à distribuição *a posteriori* da proporção de itens conformes (p). A resolução do problema será permeada pela utilização das técnicas de metaheurísticas híbridas inteligentes.

5.1. SISTEMAS HÍBRIDOS INTELIGENTES

Os seres humanos podem ser considerados um bom exemplo de “máquinas” que possuem, e transmitem informações de forma híbrida. As atitudes e ações dos seres humanos são governadas pela combinação de informações genéticas e aquelas adquiridas por meio da aprendizagem. Na informação genética, chamada de genótipo, as características que nascem com o indivíduo, em forma de codificação genética, são herdadas dos pais. O fenótipo, por outro lado, é a combinação das potencialidades fornecidas pelo genótipo com as influências do meio ambiente.

Além do mais, são as informações dos nossos genes que garantem, sobretudo, o sucesso da nossa sobrevivência. Essa sobrevivência que tem sido testada durante milhões de anos de evolução.

O aprendizado humano consiste de uma variedade de processos complexos que utilizam informações adquiridas a partir de interações com o meio externo. É a combinação desses diferentes tipos de métodos de processamento de informações que tem habilitado a raça humana a ter sucesso na sua sobrevivência em meios dinâmicos que se modificam a todo instante.

Mais especificamente, essa forma completa e complexa de processamento de informação híbrida dos seres humanos está sendo replicada para a geração de máquinas que, por sua vez, se adaptam e estão cada vez mais sendo controladas por núcleos computacionais inteligentes. Essa inteligência, denominada de inteligência artificial (IA), tem produzido resultados motivadores em resoluções de algumas tarefas particulares. Mas certos problemas complexos não podem ser resolvidos satisfatoriamente por uma única técnica inteligente, pois tais técnicas possuem pontos fracos e fortes, sendo adequadas para alguns problemas e não para outros. Essas limitações têm sido a grande motivação para o estudo de sistemas híbridos inteligentes (SHI), em que duas ou mais técnicas de IA são combinadas com o intuito de superar as respectivas limitações de técnicas individuais.

Os SHI também ganham ainda mais destaque quando se considera a gama de aplicações em que tais metodologias podem ser implementadas. Em geral, os domínios complexos têm muitas componentes de problemas diferentes, cada qual podendo requerer um tipo específico de processamento. Além do mais, os SHI podem ser combinados com diferentes técnicas, inclusive com sistemas convencionais de computação, como os próprios programas da microsoftware WindowsTM.

Os motivos para a criação das metaheurísticas híbridas são inúmeros, mas se pode resumir a três:

- ***Técnicas de intensificação:*** Integração de, pelo menos, duas técnicas heurísticas diferentes em que se deseja compensar o ponto fraco de uma técnica com o ponto forte da outra.

- **Multiplidade de aplicações em tarefas:** Um sistema híbrido inteligente é criado tendo em vista que uma única técnica não é capaz de ser aplicável aos muitos subproblemas que certa situação prática possa apresentar.
- **Realização de múltipla funcionalidade:** A criação de sistemas híbridos que exibam a capacidade de processamento múltiplo de informações dentro de uma arquitetura de eventos. Esses sistemas imitam ou emulam diferentes técnicas de processamento.

Constata-se que são muitas as técnicas e ferramentas direcionadas à resolução de problemas de otimização multiobjetivo. Em especial, têm surgido nos últimos anos uma vontade e uma predisposição muito significativa dos pesquisadores em combinar ou mesclar duas ou mais técnicas de inteligência artificial para a criação de sistemas híbridos inteligentes.

Em vista disso, a segunda parte dessa tese utilizará essa mesma linha de pensamento e combinará os métodos heurísticos de algoritmos genéticos com redes neurais e métodos de busca multiobjetivo para resolver alguns problemas de interesses inferenciais sobre a distribuição marginal *a posteriori* da proporção de itens conformes, discutido na seção 4.3.1.

5.2. O PROBLEMA DE OTIMIZAÇÃO MULTIOBJETIVO

5.2.1. DESCRIÇÃO DO PROBLEMA

Na seção 4.3 foi apresentada a construção da função de verossimilhança ordenada (expressão 4.4) que dependia basicamente de um determinado critério a ($1 \leq a \leq m$) de classificação final de cada item, no caso usado como $a = m/2$, após as m classificações e de quatro outras quantidades: n_C , $n_{\bar{C}}$, C e \bar{C} .

Assim, considere que, sobre um espaço X de todos os resultados possíveis do experimento, exista um vetor de funções dos dados amostrais constituído por $\mathbf{x} = (n_C, n_{\bar{C}}, C, \bar{C})$ e que exista ainda um vetor de parâmetros (p, e_1, e_2) sobre o hipercubo $[0;1]^3$. Além disso, seja $f(p, e_1, e_2)$ a distribuição conjunta *a priori* da tripla citada e $f(\mathbf{x} | p, e_1, e_2)$ a função de verossimilhança dos parâmetros (p, e_1, e_2) considerando os dados \mathbf{x} (equação 4.4). Logo, a distribuição *a posteriori* conjunta de (p, e_1, e_2) , denotada por $f(p, e_1, e_2 | \mathbf{x})$, dado o vetor \mathbf{x} é proporcional ao produto da distribuição *a priori* e a função de verossimilhança e está descrita em (5.1).

$$f(p, e_1, e_2 | \mathbf{x}) \propto f(\mathbf{x} | p, e_1, e_2) f(p, e_1, e_2) \quad (5.1)$$

em que $f(\mathbf{x} | p, e_1, e_2) = P(C = c, \bar{C} = \bar{c}, n_C = k | p, e_1, e_2)$.

A distribuição marginal *a posteriori* de p é a integração dupla sobre os parâmetros e_1 e e_2 , calculada como

$$f(p | \mathbf{x}) = \int_0^1 \int_0^1 f(p, e_1, e_2 | \mathbf{x}) de_1 de_2 \quad (5.2)$$

Portanto, o outro interesse principal deste capítulo é estudar as propriedades inferenciais e preditivas dos estimadores da proporção de itens conformes (p). Com esse intuito, apesar de não aparecer explicitamente na expressão 5.2 e nem na definição do vetor de dados \mathbf{x} , o tamanho da amostra e o número de classificações repetidas configuram informações fundamentais para obter as quantidades amostrais n_C , $n_{\bar{C}}$, C e \bar{C} .

Além do mais, sobre a distribuição marginal *a posteriori* de p ainda residem dois interesses principais: estudar a sua variância *a posteriori* e a amplitude do intervalo de credibilidade *a posteriori* para um nível de credibilidade $100(1-\alpha)\%$.

Especificamente, a variância *a posteriori* de p ($\hat{\sigma}_p^2$) pode ser calculada por meio da expressão

$$\hat{\sigma}_p^2 = \int_0^1 (p - E(p | \mathbf{x}, n, m, a))^2 f(p | \mathbf{x}, n, m, a) dp \quad (5.3)$$

em que $E(p | \mathbf{x}, n, m, a) = \int_0^1 pf(p | \mathbf{x}, n, m, a) dp$ é o seu valor médio.

Sob a ótica da teoria de otimização, um problema que temos que resolver é encontrar os menores valores de n , m e o critério de decisão a ($1 \leq a \leq m$) que, no pior caso, limite a máxima variância *a posteriori* de p abaixo de uma variância σ^2 preestabelecida para todos os resultados possíveis do experimento, ou seja

$$\max_{\mathbf{x} \in X} \{ \hat{\sigma}_p^2 \} \leq \sigma^2 \quad (5.4)$$

Para deixar claro sobre a intratabilidade do problema, a TABELA 5.1 apresenta o total de experimentos possíveis que pode ser criado a partir da tripla (n, m, a) , considerando todos os possíveis valores de a , n_C , $n_{\bar{C}}$, C e \bar{C} .

TABELA 5.1: Total de experimentos possíveis que podem ser gerados a partir da definição dos parâmetros de entrada n , m e a

n	m	a	Total de Experimentos Possíveis
10	3	1, ..., 3	1.188
10	5	1, ..., 5	6.105
10	7	1, ..., 7	17.402
50	3	1, ..., 3	111.928
50	5	1, ..., 5	650.505
50	7	1, ..., 7	1.948.982
100	3	1, ..., 3	863.853
100	5	1, ..., 5	5.101.005
100	7	1, ..., 7	15.377.957
300	3	1, ..., 3	22.771.553
300	5	1, ..., 5	135.903.005
300	7	1, ..., 7	411.393.857
600	3	1, ..., 3	181.083.103
600	5	1, ..., 5	1.083.606.005
600	7	1, ..., 7	3.283.567.707

Observe que, se o objetivo fosse encontrar os menores n e m que atendessem à expressão 5.4 por meio de um procedimento de busca exaustiva, seria necessário pesquisar mais de 3 bilhões de experimentos se o horizonte de busca estivesse limitado à $n=600$ e $m=7$ (TABELA 5.1). Se cada experimento consumisse 1 segundo, seria preciso aproximadamente 105 anos de esforço computacional, em termos de tempo, inviabilizando assim a solução. Por isso, será imprescindível o uso de técnicas especializadas e inteligentes na busca da solução. Esta constatação é válida para o caso de busca de um único critério de decisão (variância *a posteriori* ou amplitude do intervalo de credibilidade) tornando-se ainda mais relevante no caso multiobjetivo.

Além disso, como já discutido na seção 4.4, outro interesse relacionado ao estudo da distribuição marginal *a posteriori* de p é a determinação do seu intervalo de credibilidade. Portanto, o objetivo principal sobre esse tema é encontrar um intervalo MDP (máxima densidade de

probabilidade) com probabilidade $1-\alpha$ ($0<\alpha<1$) com a maior amplitude possível, por exemplo $w'(\mathbf{x}, n, m, a)$, que atingisse, no máximo, uma amplitude w preestabelecida, de modo que

$$w'(\mathbf{x}, n, m, a) \leq w \quad (5.5)$$

Observe que, em geral, $w'(\cdot)$ dependerá do tamanho amostral (n), do número de classificações repetidas (m) e do critério de classificação final ($1 \leq a \leq m$). Novamente, trata-se de investigar os valores ótimos de n , m e a sob a ótica do pior cenário possível.

Joseph, Wolfson e Du Berger (1995) estudaram a obtenção de tamanhos de amostra em problemas de dados dicotômicos. Segundo os autores, se p for unidimensional e $f(p|x)$, que depende de n , for unimodal, então um intervalo $[a, b]$ será MDP se e somente se $p_1 \neq p_2$ com $f(p_1 | x) \geq f(p_2 | x)$ para todo $p_1 \in [a, b]$ e para todo $p_2 \notin [a, b]$, isto é, se a função densidade a *posteriori*, para todo p_1 dentro do intervalo MDP, for maior do que qualquer outra densidade para todo p_2 fora do intervalo MDP. Os autores ainda demonstraram três critérios bayesianos para determinação do tamanho da amostra.

O primeiro critério refere-se à cobertura média do intervalo de credibilidade para a distribuição *a posteriori* de p . Esse critério garante que, em média, a probabilidade de cobertura da distribuição *a posteriori* é, no mínimo, $1-\alpha$. Dessa forma, para uma dada amplitude de intervalo MDP fixo w , o menor tamanho amostral n , cuja probabilidade de cobertura é, no mínimo, $1-\alpha$ deve satisfazer à expressão

$$\int_x \left\{ \int_{h_1}^{h_2} f(p | \mathbf{x}, n, m, a) dp \right\} f(\mathbf{x}) dx \geq 1 - \alpha \quad (5.6)$$

em que $h_1 = h(x, n, m, a)$ é o limite inferior do intervalo de credibilidade que depende dos dados x e dos parâmetros n, m e a e $h_2 = h_1 + w'(x, n, m, a)$ é o limite superior do intervalo de credibilidade.

Além disso, $f(x) = \int_0^1 \int_0^1 \int_0^1 f(x | p, e_1, e_2) f(p, e_1, e_2) dp de_1 de_2$ é a distribuição marginal de x e $f(p | x, n, m, a)$ é a densidade *a posteriori* de p dada pela equação 5.2.

Com um raciocínio análogo, o segundo critério sugere uma maneira alternativa de selecionar o tamanho da amostra baseado numa probabilidade de cobertura $1-\alpha$ fixa para o intervalo de credibilidade MDP da proporção de itens conformes.

Assim, considere um intervalo MDP para p com amplitude w e uma probabilidade de cobertura $1-\alpha$, ambos fixos. Suponha, ainda, que cada resultado x do experimento exija um comprimento $w'(x, n, m, a)$ de intervalo para atingir a cobertura desejada. Desse modo, a estruturação desse método garante que, para algum $x \in X$,

$$\int_{h_1}^{h_2} f(p | x, n, m, a) dp = 1 - \alpha \quad (5.7)$$

E o problema se resume, então, em selecionar n, m e a , tais que a esperança (com respeito a x) das amplitudes desses intervalos de tamanho $w'(x, n, m, a)$ seja, no máximo, w , ou seja

$$\int_X w'(x, n, m, a) f(x) dx \leq w \quad (5.8)$$

Mas, apesar de bastante intuitivo e direto, o método do tamanho médio do intervalo pode ser criticado no sentido de garantir, apenas na média, a amplitude do intervalo desejado w . Portanto, o procedimento não fornece nenhuma garantia para algum conjunto de dados x em particular.

Por isso, uma abordagem mais conservativa que garanta, ao mesmo tempo, a amplitude do intervalo MDP e a probabilidade de cobertura $1-\alpha$ sobre todos os possíveis vetores x deve ser definida. Dessa forma, o terceiro procedimento, conhecido como critério do pior resultado, busca os valores de n , m e a que satisfaçam a desigualdade

$$\inf_{x \in X} \left\{ \int_{h_1}^{h_1+w} f(p | x, n, m, a) dp \right\} \geq 1 - \alpha \quad (5.9)$$

em que w e α são valores prefixados. Essa especificação é similar ao critério que garante que a máxima variância *a posteriori* não excede um valor predefinido (5.4), pois ambas são construídas sob a suposição do pior cenário possível.

Assim, diante de tudo o que foi discutido e dentro do âmbito de otimização multiobjetivo, a principal meta é encontrar os valores ótimos dos parâmetros n , m e a que garantam, simultaneamente, tanto a variância *a posteriori* de p (5.4) quanto a amplitude do intervalo MDP $w'(x, n, m, a)$, de modo que esse tenha uma probabilidade mínima de cobertura de $1-\alpha$ (baseado na expressão 5.9), tal que de certos não excedam os valores preestabelecidos. Em outras palavras, uma vez definidas a variância σ^2 máxima *a posteriori* e a amplitude máxima w do intervalo MDP, buscase o vetor (n, m, a) que atinja, no máximo, os seus respectivos valores especificados.

Portanto, nota-se que o problema de otimização objeto de estudo não é de fácil solução (ver TABELA 5.1), pois retrata um processo de otimização com restrições tanto nos parâmetros de entrada quanto nas funções objetivos. Exigirá um enorme esforço computacional, uma vez que as medidas desconhecidas no espaço das variáveis (n, m, a) em conjunto com os resultados experimentais possíveis geram uma gama grande de verificações. A apresentação da metodologia proposta basear-se-á na definição de ferramentas analíticas utilizadas no processo de otimização.

Inicialmente o campo de verificações das soluções factíveis será reduzido por meio de uma definição matricial, chamada aqui de Matriz-Trabalho (MT). Em seguida, esse processo restrito de busca será desconsiderado, de modo a permitir a construção de um processo de otimização mais preciso.

5.2.2. MATRIZ-TRABALHO

Com objetivo de reduzir a busca por todos os resultados experimentais possíveis, define-se como Matriz-Trabalho (MT) o número total de pontos avaliados num determinado experimento, provenientes da multiplicação do número de itens avaliados (n) e do número de classificações repetidas (m) realizadas, ou seja

$$MT(n; m) = n \times m \quad (5.10)$$

Desse modo, investigar as propriedades inferenciais de experimentos com Matrizes-Trabalho de tamanho R ($MT(n; m) = R$) é garantir que os parâmetros ótimos obtidos representarão sempre conjuntos de pontos com o mesmo padrão (trabalho) de esforço computacional. Essa definição foi baseada na seguinte questão: é melhor planejar um experimento com $n=6$ e $m=1$ ou com $n=3$ e $m=2$? Baseado na Matriz-Trabalho, essas duas situações são equivalentes, pois representam a mesma Matriz-Trabalho. Evidentemente, o objetivo maior sobre a questão formulada será, sobretudo, identificar qual a melhor combinação (n, m), considerando o mesmo esforço computacional, que satisfaça os critérios especificados da variância e amplitude de intervalo obtidos *a posteriori*.

Assim, a busca pelos valores ótimos de n e m ficará restrita a todos os pares de divisores (n, m) que satisfazem a condição de que $MT(n; m) = R$. Por exemplo, para uma Matriz-Trabalho de tamanho 30 ($R=30$), se $n=1$ implica em $m=30$; se $n=2$ implica em $m=15$; se $n=3$ implica em $m=10$; e, para um caso geral, se $n=\vartheta$ implica em $m=R/\vartheta$, sendo ϑ um divisor de R .

Assim, para uma Matriz-Trabalho fixa, o conhecimento do tamanho amostral determina automaticamente o número de classificações repetidas, tornando esse último parâmetro determinístico no processo de otimização. O ideal, portanto, num processo de otimização, é tentar definir uma Matriz-Trabalho que contenha muitos divisores, possibilitando assim, um conjunto maior de possibilidades no espaço de busca das soluções factíveis, ou seja, das soluções que atendem a todas as restrições do problema de otimização.

Por fim, pode-se ressaltar que a Matriz-Trabalho é, na verdade, uma estratégia para contornar o problema de ter que pesquisar todas as combinações (n, m) possíveis. Conforme os resultados já apresentados na TABELA 5.1, constatou-se que é computacionalmente inviável investigar todos os pares (n, m) .

5.2.3. FUNÇÃO OBJETIVO

Um problema geral de otimização multiobjetivo consiste em encontrar um vetor de variáveis de decisão (solução) que satisfaça às restrições estabelecidas e, concomitantemente, otimize uma função vetorial, cujos elementos correspondem às funções objetivos. Essas funções representam os critérios de otimalidade do problema em estudo. Portanto, “otimizar” significa encontrar soluções com todos os valores dos objetivos que não podem ser melhorados simultaneamente. Além disso, como o processo de maximização pode ser transformado em problema de minimização, bastando apenas multiplicar a função por -1 (princípio da dualidade). É comum reescrever tais problemas somente em termos de maximização.

Formalmente, isso pode ser definido da seguinte maneira:

$$\text{Maximizar } \mathbf{z} = \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_r(\mathbf{x}))$$

$$\text{Sujeito a } \mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_s(\mathbf{x})) \leq \mathbf{k}$$

$$\mathbf{x} = (x_1, x_2, \dots, x_u) \in \mathbf{X}$$

$$\mathbf{z} = (z_1, z_2, \dots, z_r) \in \mathbf{Z}$$

em que \mathbf{x} é o vetor de decisão, \mathbf{z} é o vetor objetivo, \mathbf{X} denota o *espaço de decisões* e $\mathbf{Z} = \mathbf{f}(\mathbf{X})$ é a imagem de \mathbf{X} , denominada de *espaço objetivo*. Adicionalmente, as restrições $\mathbf{g}(\mathbf{x}) \leq \mathbf{k}$, $\mathbf{k} \in \mathfrak{R}^s$ e o espaço \mathbf{X} determinam o conjunto de soluções factíveis, ou seja, uma determinada solução \mathbf{x} só será considerada factível se ela pertencer ao espaço de decisões e atender a todas as restrições $\mathbf{g}(\mathbf{x})$.

No entanto, quais são os valores ótimos para os parâmetros (n, m, a) que maximizem, simultaneamente, a amplitude média do intervalo de credibilidade e a variância *a posteriori* de p , dada a uma Matriz-Trabalho de tamanho R , a uma variância máxima *a posteriori* σ^2 e a um comprimento máximo de intervalo w de cobertura $1-\alpha$, todos predefinidos? Especificamente, buscam-se os valores dos parâmetros de entrada que satisfazem o pior cenário possível, em termos de variabilidade *a posteriori* e de comprimento médio de intervalo de credibilidade de máxima densidade. E esse problema pode ser representado pela seguinte estrutura básica

$$\text{Maximizar } \mathbf{z} = \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})) \quad (5.11)$$

$$\text{Sujeito a } \max_{x \in X} \{f_1(\mathbf{x})\} \leq \sigma^2;$$

$$\inf_{x \in X} \left\{ \int_{h_1}^{h_1+w} f(p|x) dp \right\} \geq 1 - \alpha;$$

$$\text{MT}(n; m) = \mathbf{R}; n \in \mathbf{Z}^+; m \in \mathbf{Z}^+; a \in \mathbf{Z}^+ \text{ e } 1 \leq a \leq m;$$

$$\sigma^2 > 0; w > 0; 0 < \alpha < 1;$$

$\mathbf{x} \in \mathbf{X} \rightarrow$ Resultados Experimentais;

$(n, m, a) \rightarrow$ Variáveis de Decisão;

em que a função objetivo 1, $f_1(x) = \hat{\sigma}_p^2$, representa a variância *a posteriori* de p e a função objetivo 2, $f_2(x)$, é a amplitude do intervalo de credibilidade MDP da distribuição marginal *a posteriori* de p com cobertura $1-\alpha$ e $f(p|x)$ é a distribuição *a posteriori* de p , respectivamente dados em (5.4) e (5.9).

Um grau de complicação para o cenário de otimização apresentado é o fato de que os valores ótimos deverão ser otimizados em espaços de busca de valores inteiros, o que dificulta a verificação dos critérios de parada do processo otimizador.

Tendo em vista que o problema de otimização discutido está inserido na classe de técnicas de otimização combinatória, é de fundamental importância o uso de metodologias adequadas e adaptadas para esse fim. O representante que tem sido extensamente empregado ao longo das últimas duas décadas são os algoritmos genéticos. Essas estruturas computacionais inteligentes e versáteis têm fornecido significativos resultados em todas as áreas do conhecimento e será utilizado,

no sentido de obter resultados mais rápidos e precisos. Será discutida na seção 5.3 uma versão híbrida de um algoritmo genético inteiramente dedicado à otimização multiobjetivo.

5.3. ALGORITMO GENÉTICO HÍBRIDO MULTIOBJETIVO PROPOSTO

Os algoritmos genéticos (AG) foram inicialmente introduzidos por Holland (1975) e fazem parte da área de computação evolutiva, que, por sua vez, constituem uma classe de métodos estocásticos computacionais inspirados na evolução natural das espécies. Os AGs são métodos robustos, flexíveis e têm a capacidade de produzir boas soluções para problemas complexos e de grande porte em um tempo computacionalmente viável. Por essas razões têm sido aplicados com enorme sucesso em uma grande variedade de situações práticas que envolvem otimização combinatória.

Os algoritmos genéticos constituem-se em uma meta-heurística e conseqüentemente não são ferramentas computacionais perfeitas e não existe uma garantia de que a solução apresentada seja ótima apesar de que muito provavelmente deverá ser uma boa solução. Por isso, tem sido muito comum hoje a construção de algoritmos evolutivos mais adaptados e mais específicos para resolver determinadas situações com maior precisão. Essa personalização tem sido conseguida com a hibridização, ou seja, com a combinação de dois ou mais métodos de heurísticas de otimização, no sentido de criar um outro algoritmo mais eficiente.

De um modo geral, um AG é composto, basicamente, por um conjunto de tarefas que são desenvolvidas antes, durante e depois da construção das gerações evolutivas. E são essas tarefas ou procedimentos (FIGURA 5.1) que serão discutidos como uma proposta de um algoritmo genético híbrido multiobjetivo (AGHM), enfocando sempre o seu uso para a solução dos problemas de

otimização que envolvem a obtenção da tripla (n, m, a) a partir de especificações relativas à distribuição marginal *a posteriori* de p .

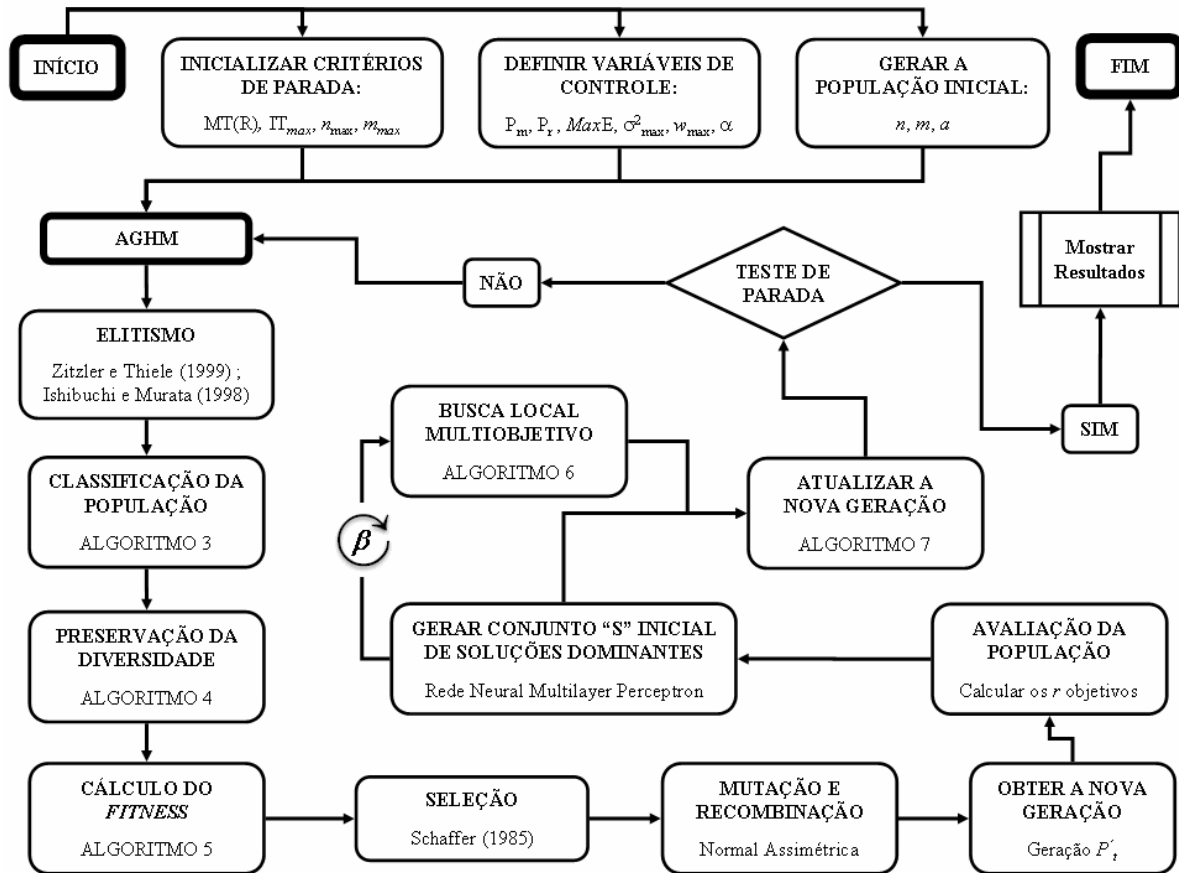


FIGURA 5.1: Tarefas realizadas para o cálculo do vetor ótimo (n, m, a)

5.3.1. DEFINIÇÃO DO CROMOSSOMO

No algoritmo genético, um indivíduo, que é uma possível solução para o problema multiobjetivo, é basicamente descrito por um cromossomo, normalmente representado por uma estrutura de dados qualquer. Os cromossomos possuem todas as informações do indivíduo e são compostos por genes, que nada mais são do que as características dos indivíduos. Além disso, cada

gene é um elemento do vetor cromossomo que, tipicamente, recebe um valor numérico, um caractere ou uma cadeia de caracteres escrita em algum código e que expressa alguma característica do indivíduo.

Finalmente, cada parâmetro do vetor de entradas do AGHM está associado a um gene, e para cada gene é definido um intervalo ou domínio dos possíveis valores. No caso específico do problema investigado nesse documento, o cromossomo, de codificação real, é definido pelos seguintes elementos:

$$\Psi = [\phi_1, \phi_2, \phi_3] = [n, m, a] \quad (5.12)$$

em que o domínio de cada um dos parâmetros de entrada já foi definido na expressão 5.11.

Assim, todo o desenvolvimento do AG, incluindo os operadores genéticos, será baseado no vetor Ψ que representa um indivíduo ou uma solução para o problema. Mas, para o AGHM proposto, decidiu-se combinar as seguintes estratégias, a fim de realizar uma busca mais eficaz e encontrar várias soluções em paralelo, quais sejam:

- Elitismo: é mantido um subconjunto de soluções dominantes encontradas até o momento.
- Classificação da população de acordo com a dominância das soluções para calcular a função de aptidão (*fitness*) de cada solução.
- Preservação de diversidade na população.
- Uso de busca local multiobjetivo baseada na relação de dominância de Pareto.

5.3.2. ESTRATÉGIA DO ELITISMO

Segundo Zitzler e Thiele (1999), um conjunto de pontos eficientes é uma região, gerada no espaço das soluções factíveis do problema investigado, em que todas as soluções são equivalentemente plausíveis, isto é, não existem soluções que melhorem todos os objetivos simultaneamente. Esses pontos são, normalmente, chamados de soluções não-dominadas (FIGURA 5.2 para um problema de minimização de f_1 e f_2) e a união dessas soluções forma o que se costuma denominar de fronteira de Pareto ou conjunto Pareto-ótimo (FIGURA 5.3).

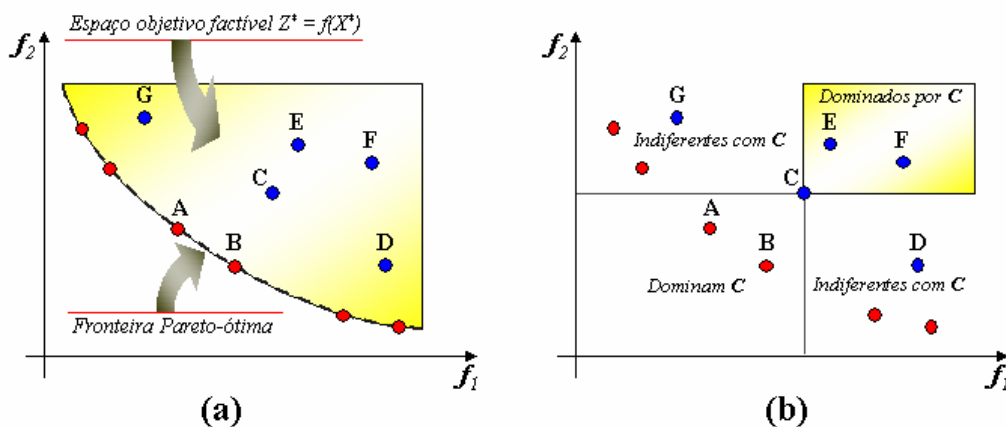


FIGURA 5.2: Exemplo fictício da relação de dominância

A incorporação da estratégia de elitismo em algoritmos genéticos multiobjetivos é mais complexa que em algoritmos genéticos mono-objetivos. Em algoritmos genéticos mono-objetivos, geralmente, a melhor solução é mantida na população. No caso multiobjetivo, no lugar da melhor solução, existe um conjunto de soluções dominantes (melhor aproximação do conjunto Pareto-ótimo) cujo tamanho pode crescer de acordo com o número de objetivos do problema. Por isto, é necessário fixar critérios para a escolha das soluções de elite que serão adicionados à população seguinte.

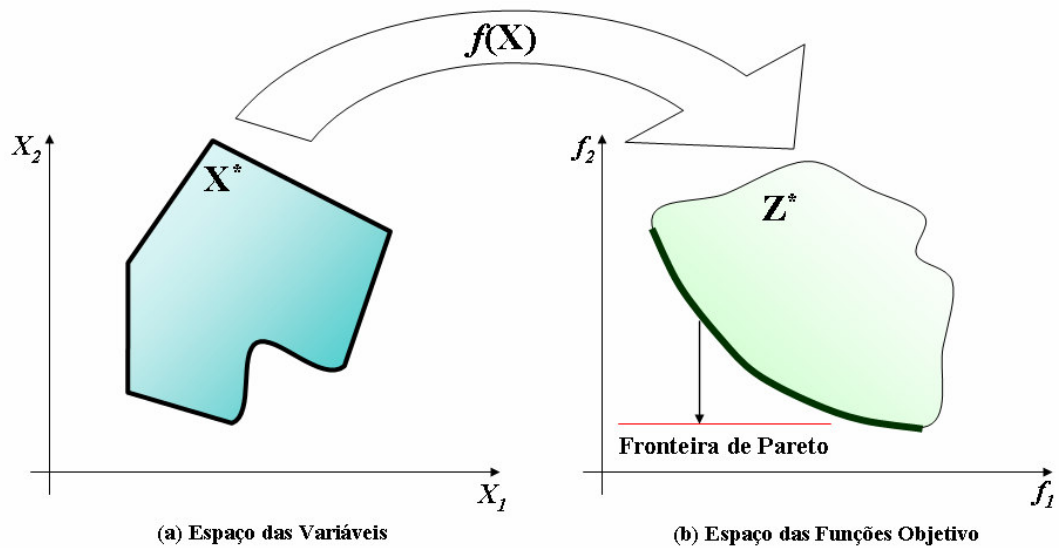


FIGURA 5.3: Relação entre os espaços factíveis das variáveis e dos objetivos

Na metaheurística AGHM proposta, é usada uma estratégia de elitismo que combina as estratégias discutidas por Zitzler e Thiele (1999) e por Ishibuchi e Murata (1998). Assim, selecionam-se N_{elite} soluções dominantes encontradas até a iteração atual t para serem adicionadas à população seguinte P_{t+1} . Então, na iteração $t+1$ constrói-se uma população combinada $P_{t+1} \cup E_t$, em que E_t é o conjunto das N_{elite} soluções de elite selecionadas a partir do conjunto de soluções dominantes D_t encontradas até a iteração t . A seleção das N_{elite} soluções de elite pode ser baseada em um critério que garanta a escolha de soluções dominantes dispersas, isto é, pontos no espaço objetivo distribuídos sobre toda a fronteira dominante atual.

Na implementação do algoritmo proposto, essas soluções de elite são selecionadas aleatoriamente segundo uma distribuição uniforme. A cada iteração, escolhe-se aleatoriamente N_{elite} soluções de elite do conjunto dominante. A razão para a não utilização de uma técnica que garanta a distribuição das soluções é: se o conjunto de soluções dominantes D não é alterado de uma iteração para outra, uma técnica determinística, como por exemplo, a técnica de agrupamento (MORSE,

1980), de uma iteração para outra, sempre selecionará as mesmas N_{elite} soluções dispersas (centróides). Uma possibilidade mais custosa não explorada aqui consiste em determinar os grupos e escolher um ponto aleatório de cada grupo.

5.3.3. CLASSIFICAÇÃO DA POPULAÇÃO

Cada solução é avaliada por uma função que retrata o potencial dessa solução (desse indivíduo) ser selecionado para participar das gerações futuras. A essa função dá-se o nome de função de ajuste ou *fitness*. Para atribuir um valor de *fitness* às soluções no AGHM, primeiramente classificam-se as soluções da população combinada $P_t \cup E_{t-1}$, em que P_t é a população da iteração atual e E_{t-1} é o conjunto de soluções de elite encontrado até a iteração anterior. Adota-se uma técnica de ordenação por dominância (*nondominated sorting technique*), inicialmente proposta por Goldberg (1989) e utilizada por Srinivas e Deb (1995) e Deb *et al.* (2000).

Essa técnica ordena a população de acordo com a relação de dominância das soluções, construindo uma população classificada em K conjuntos dominantes F_1, F_2, \dots, F_K . Em cada conjunto F_i , todas as soluções são indiferentes (ou seja, ninguém domina ninguém). A FIGURA 5.4 mostra um exemplo para o caso de um problema de minimização bi-objetivo em que a população de pontos $f(P_t \cup E_{t-1})$ no espaço objetivo é classificada em $K=3$ fronteiras $f(F_1)$, $f(F_2)$ e $f(F_3)$. Observe que os pontos na fronteira $f(F_1)$ são os dominantes (correspondem às soluções mais aptas) e os pontos em $f(F_2)$ dominam os pontos em $f(F_3)$.

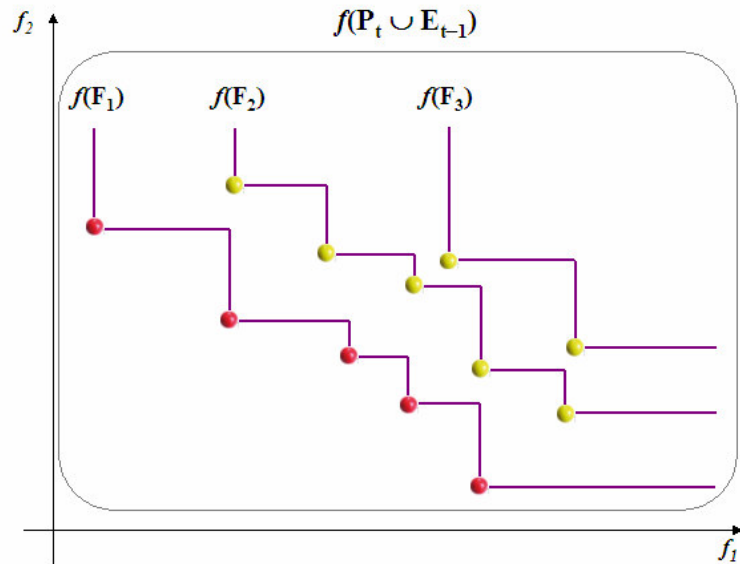


FIGURA 5.4: Exemplo do método de classificação dos conjuntos de soluções de acordo com o valor do *fitness*

Deb *et al.* (2000) propuseram um algoritmo para classificar uma população P de soluções (considerando os vetores objetivos) implementado em tempo $O(r \times N^2)$, em que N é o tamanho da população a ser ordenada e r é o número de objetivos. Esse algoritmo é denominado *fast nondominated sort* e é uma versão melhorada da ordenação usada no algoritmo genético de Srinivas e Deb (1995) implementado em tempo $O(r \times N^3)$. Nesse algoritmo, para cada solução x de uma população P determinam-se:

- n_x : o número de soluções que dominam x ;
- S_x : o conjunto de soluções dominadas por x ;

Em seguida, identificam-se todas as soluções com $n_x = 0$ (soluções não dominadas), e essas são armazenadas em um conjunto F_1 , que constitui a fronteira atual. Para cada solução y dominada por alguma solução x da fronteira atual, o valor de n_y é diminuído em uma unidade ($n_y = n_y - 1$). Então, as soluções y que tenham $n_y = 0$ são armazenadas num novo conjunto F_2 (fronteira atual). Esse processo continua até que todas as soluções possuam $n_y = 0$, isto é, até que todas as soluções sejam armazenadas numa fronteira. A FIGURA 5.5 apresenta o pseudocódigo do algoritmo *fast nondominated sort* aplicado a uma população P .

ALGORITMO 3: FAST NONDOMINATED SORT

Entrada: P (uma população de soluções)

Saída: F_1, F_2, \dots, F_K (população classificada em K sub-populações)

1) Para cada $x \in P$ faça

 Para cada $y \in P$ faça

 Se x domina y então $S_x = S_x \times \{y\}$ (S_x é o conjunto de soluções dominadas por x).

 Senão, se y domina x então $n_x = n_x + 1$.

 Se $n_x = 0$ (n_x é o número de soluções que dominam x) então $F_1 = F_1 \cup \{x\}$.

 Faça $j = 1$.

2) Enquanto $F_j \neq \emptyset$ faça

 Faça $F_{j+1} = \emptyset$.

 Para cada $x \in F_j$ faça

 Para cada $y \in S_x$ faça

 Faça $n_y = n_y - 1$. Se $n_y = 0$ então $F_{j+1} = F_{j+1} \cup \{y\}$.

 Faça $j = j + 1$.

FIGURA 5.5: Algoritmo *Fast Nondominated Sort* criado por Deb *et al.* (2000)

Observe que a primeira fase do algoritmo *fast nondominated sort* requer tempo $O(r \times N^2)$, pois as soluções da população são comparadas duas a duas, e para cada comparação de duas soluções, comparam-se os r objetivos. Enquanto que a segunda fase do algoritmo, no pior caso (quando cada fronteira contém um único ponto), requer tempo $O(N^2)$. Portanto, a complexidade do algoritmo é $O(r \times N^2) + O(N^2) = O(r \times N^2)$. Na metaheurística AGHM proposta é usado o algoritmo *fast nondominated sort* para classificar a população combinada $P_t \cup E_{t-1}$.

5.3.4. PRESERVAÇÃO DA DIVERSIDADE DA POPULAÇÃO E CÁLCULO DO *FITNESS*

Depois de classificar a população $P_t \cup E_{t-1}$ em K conjuntos dominantes F_1, F_2, \dots, F_K , deve-se definir e utilizar uma função que represente o grau de aptidão (*fitness*) que uma determinada solução tem de continuar no processo evolutivo de reprodução. Normalmente, a própria função a ser otimizada é utilizada como função de aptidão ou como parte integrante dessa. Todavia, nesse trabalho, apenas a variância *a posteriori* de p (5.3) será usada como medida de ajuste, no sentido de que valores menores dessa indicarão aqueles indivíduos (soluções) mais aptos.

Adicionalmente, as soluções de um conjunto F_K possuirão, de fato, maior potencial de reprodução (*fitness*) que as soluções de um conjunto F_j ($j > K$). Dado que as soluções em um determinado conjunto F_K são indiferentes (possuem o mesmo grau de dominância), é necessário aplicar alguma técnica para diferenciar essas soluções. Então, na metaheurística AGHM, o cálculo do *fitness* é baseado na técnica proposta por Deb *et al.* (2000).

Essa metodologia, quando aplicada a cada conjunto F_K , tenta preservar a diversidade da população $P_t \cup E_{t-1}$. Para cada solução $x \in F_K$, determina-se uma estimativa do número de soluções localizadas ao redor de x , considerando o vetor de objetivos das soluções dada pela distância de

aglomeração (FIGURA 5.6). Para uma determinada fronteira $f(F_K)$, calcula-se a distância de aglomeração ($dist$) dos seus pontos z^1, \dots, z^{np} , em que np é o número de pontos na fronteira. Os pontos extremos z^1 e z^{np} , geralmente, terão altos valores de distância de aglomeração. Uma forma de atribuir o ajuste dos pontos de uma fronteira é "normalizar" tais distâncias pelos seus valores máximo e mínimo.

Um exemplo das distâncias de aglomeração de uma população fictícia é ilustrado na FIGURA 5.7a, no qual é obtida por meio do algoritmo padrão proposto por Deb *et al.* (2000). Paralelamente, a FIGURA 5.7b demonstra o cálculo de normalização das distâncias obtidas pelo método original. Note que, para o caso bi-objetivo, os pontos extremos de qualquer fronteira sempre possuirão a mesma distância de aglomeração é dada por $(Min_{dist} + Max_{dist})/Min_{dist}$ e $dist/Min_{dist}$ para os demais pontos dessa fronteira. Observe também que os pontos mais dispersos (ou isolados), numa fronteira, sempre possuirão maiores valores dessa medida de distância.

ALGORITMO 4: ATRIBUIÇÃO DA DISTÂNCIA DE AGLOMERAÇÃO

Entrada: $S, |S| = np$ (sub-população de tamanho np)

Saída: $dist(z), \forall z \in f(S)$ (distância de aglomeração no espaço objetivo)

- 1) Para cada ponto $z^i \in f(S)$ faça $dist(z^i) = 0$.
- 2) Para cada objetivo j faça
 - 2.1) Ordene os pontos em $f(S)$ de acordo com o objetivo j obtendo z^1, \dots, z^{np} .
 - 2.2) $dist(z^1) = v, dist(z^{np}) = v$ (v é um valor suficientemente grande)
 - 2.3) Para $i = 2$ até $(np-1)$ faça

$$dist(z^i) = dist(z^1) + (f_j(z^{i+1}) - f_j(z^{i-1})),$$
 em que $f_j(z^i)$ é o valor da j -ésima função objetivo do ponto z^i .

FIGURA 5.6: Algoritmo de cálculo da distância de aglomeração

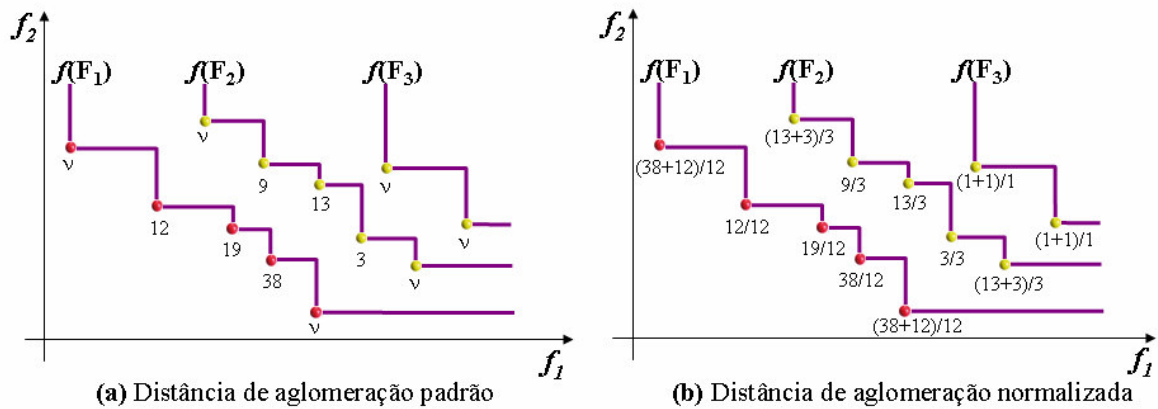


FIGURA 5.7: Cálculo das distâncias de aglomeração pelos métodos padrão e normalizado

Após obtida a distância de *aglomeração* de todos os pontos da população $P_t \cup E_{t-1}$, calcula-se o valor do ajuste (*fitness*) de cada solução por meio do Algoritmo 5 apresentado na FIGURA 5.8. Um exemplo desse procedimento foi feito utilizando as distâncias de aglomeração normalizadas ilustradas na FIGURA 5.7b que está apresentada na FIGURA 5.8.

Note que o valor do *fitness* de uma solução depende da fronteira à qual ela pertence e da sua distância de aglomeração. As soluções em F_K sempre possuirão maior potencial de reprodução que qualquer solução em F_{K+1} . Assim, o *fitness* é sempre maximizado, isto é, valores altos de *fitness* correspondem a altas probabilidades de reprodução, independente se o problema é de maximização ou de minimização.

ALGORITMO 5: CÁLCULO DO FITNESS NO AGHM

Entrada: distância de aglomeração (*crowding distance*) das K *sub-populações*

Saída: *fitness* das K *sub-populações*

1) Faça $maxfitness_{K+1} = 0$ (máximo valor de *fitness* na fronteira $K+1$).

2) Para $j = K$ diminuindo até 1 faça

2.1) Para cada solução $x \in F_j$ faça

$$fitness_x = maxfitness_{j+1} + dist(f(x))$$

2.2) $maxfitness_j = \max\{fitness(y) : y \in F_j\}$ é o máximo valor de *fitness* de F_j .

FIGURA 5.8: Algoritmo para o cálculo do *fitness* no AGHM

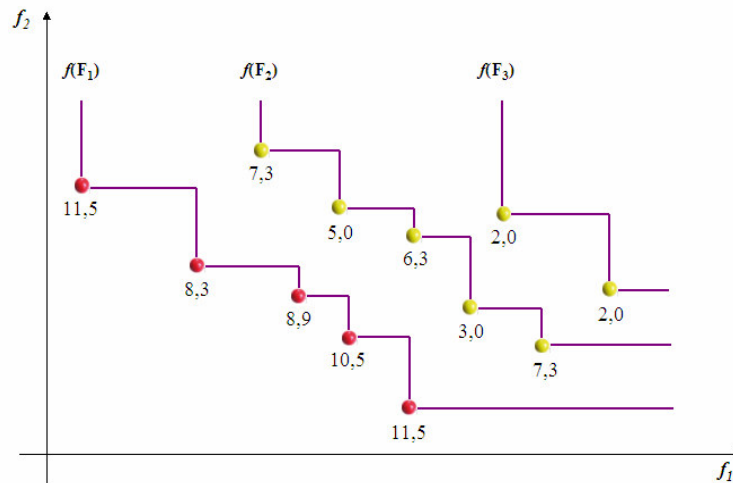


FIGURA 5.9: Cálculo do *fitness* após obtenção das distâncias de aglomeração normalizadas

Finalmente, depois calcular os valores de *fitness* de todas as soluções, a população $P_t \cup E_{t-1}$ é reproduzida como em um algoritmo genético convencional (FIGURA 5.10).

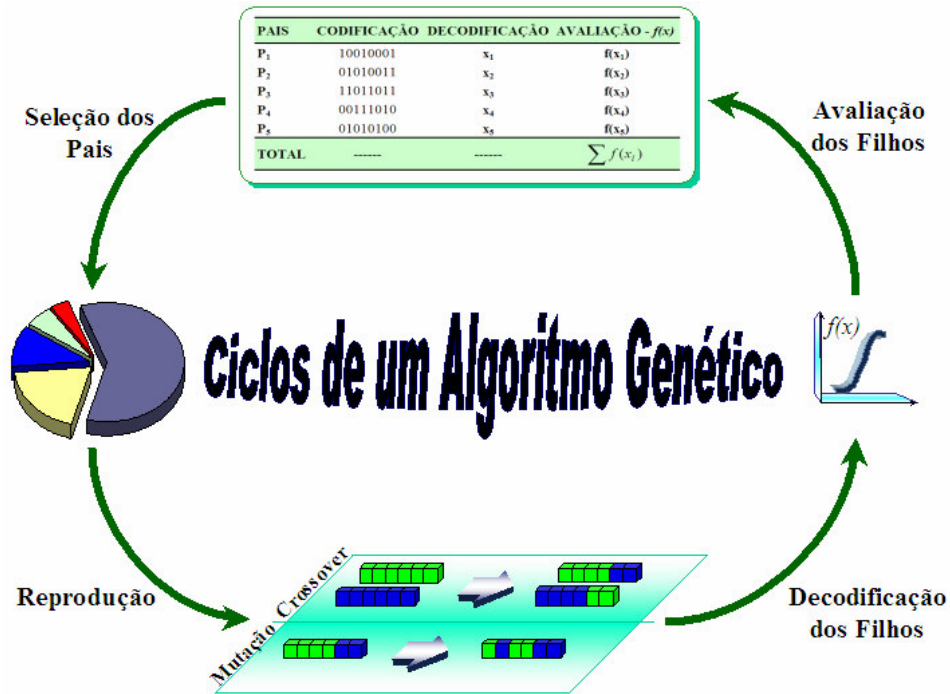


FIGURA 5.10: Estrutura geral de um algoritmo genético tradicional

5.3.5. SELEÇÃO

Depois de classificar a população combinada $P_t \cup E_{t-1}$ e calcular os valores dos *fitness* das soluções, utiliza-se o método de seleção de Schaffer (1985) para selecionar as N soluções mais aptas que formarão a população de soluções pais (*mating pool*) SP_t .

Schaffer (1985) implementou um método de seleção que consiste na separação dos objetivos. A cada iteração geram-se r subpopulações P_j , associadas aos r objetivos de tamanho N/r , em que N é o tamanho da população, e cada sub-população contém as melhores soluções, segundo um dado objetivo. Este método busca pontos segundo direções dos eixos das funções objetivos e, conseqüentemente, dá maior preferência a pontos dominantes localizados nos extremos da fronteira (FIGURA 5.11).

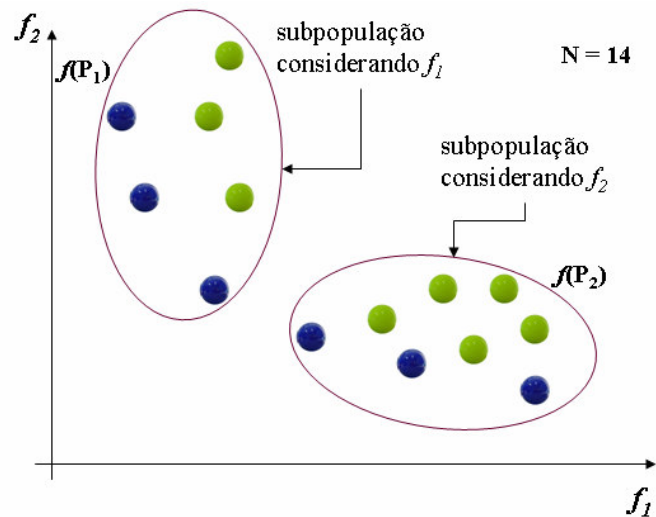


FIGURA 5.11: Exemplo do método de agrupamento de Shaffer

A partir daí, aplica-se os operadores genéticos de recombinação (*crossover*) e mutação à população SP_t construindo, assim, uma população de N novas soluções (filhos) P'_t . Tais operadores serão modificados, como pode ser visto na seção 5.3.6 para se adaptarem ao problema multiobjetivo que se deseja resolver neste trabalho.

5.3.6. MUTAÇÃO

A mutação pode ser entendida como o operador responsável pela introdução e manutenção da diversidade genética na população. Esse operador trabalha alterando, aleatoriamente, um ou mais componentes do vetor cromossomo, fornecendo, dessa forma, meios para a introdução de novos indivíduos na população. O operador mutação é aplicado aos indivíduos com uma probabilidade P_m , chamada de taxa de mutação. Michalewicz (1994) discute vários métodos de mutação para codificação real.

Mas antes de continuar a desenvolver esse tópico, convém discutir alguns detalhes sobre uma família de distribuições normais assimétricas. Essa classe de distribuições servirá de alicerce para a construção de um novo método de mutação que foi inspirado basicamente na técnica de mutação gaussiana.

Segundo Azzalini (1985), uma variável aleatória Z tem distribuição Normal Assimétrica (*skew normal*) padrão com parâmetro λ de assimetria, denotada por $Z \sim SN(0,1,\lambda)$, se a sua função densidade for dada por

$$f(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathfrak{R} \quad (5.13)$$

em que $\lambda \in \mathfrak{R}$ é uma constante, $\phi(\cdot)$ e $\Phi(\cdot)$ denotam as respectivas funções de densidade de probabilidade (fdp) e de distribuição acumulada (fda) da distribuição normal padrão $N(0, 1)$.

A FIGURA 5.12 mostra o comportamento de $f(z)$ para cinco valores de λ . A distribuição Normal padrão é um caso particular da distribuição Normal Assimétrica quando $\lambda=0$, ou seja, a família de distribuições normais assimétricas é uma classe bem mais geral e o que define seu grau de assimetria à esquerda ou à direita é o valor de λ .

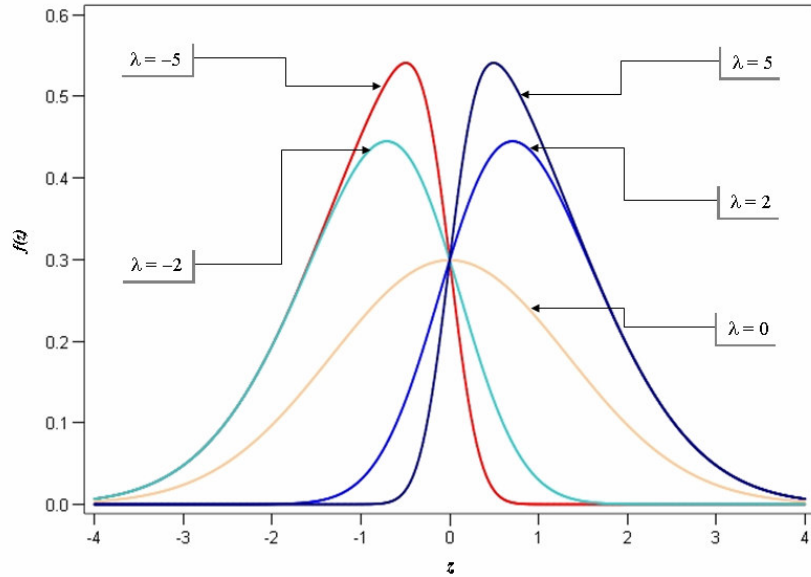


FIGURA 5.12: Função densidade da distribuição normal assimétrica para cinco valores de λ

Essa família de distribuições assimétricas tem propriedades muito interessantes. Algumas dessas são mostradas a seguir e para uma discussão mais detalhada consultar Azzalini (1985).

Propriedade 1: Se $Z \sim \text{SN}(0, 1, \lambda)$ e $Y = \mu + \sigma Z$, então $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$ com função densidade de probabilidade dada por

$$f(y) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right), \quad y \in \mathfrak{R} \quad (5.14)$$

Propriedade 2: Seja $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$, então a esperança e variância de Y são dadas por

$$E(Y) = \mu + \sigma \delta \sqrt{\frac{2}{\pi}} \quad \text{e} \quad V(Y) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2\right), \quad \text{sendo} \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \quad (5.15)$$

Assim, a distribuição normal assimétrica é bastante versátil, principalmente porque pertence à família de distribuições com parâmetros de locação e escala e possui suporte no conjunto de números reais. Foram essas características que possibilitaram a construção do operador de mutação desejado.

Após essa pequena explanação, considere o vetor cromossomo $\Psi_t = [\phi_{1t}, \phi_{2t}, \phi_{3t}]$ referente à geração t descrito pelos três componentes de entrada examinados (ver expressão 5.12). Para cada uma dessas componentes é gerado um número u de uma distribuição de Bernoulli com parâmetro P_m (probabilidade de mutação) que indicará se essa componente sofrerá ou não o processo mutacional. Caso $u=1$ para uma determinada componente k ($k \leq 3$), então o processo de mutação dessa componente será efetuado baseando-se no procedimento abaixo:

PASSO 1: Considerando LI_{ϕ_k} e LS_{ϕ_k} como os limites da componente ϕ_{kt} na geração t , obter a média ($\mu_{\phi_{kt}}$), o desvio padrão ($\sigma_{\phi_{kt}}$) e o ponto médio do intervalo ($x_{\phi_{kt}}$), respectivamente, como segue:

$$\begin{aligned}\mu_{\phi_{kt}} &= \phi_{kt} \\ \sigma_{\phi_{kt}} &= \frac{LS_{\phi_k} - LI_{\phi_k}}{6} \\ x_{\phi_{kt}} &= \frac{LI_{\phi_k} + LS_{\phi_k}}{2}\end{aligned}\tag{5.16}$$

PASSO 2: Defina uma variável aleatória Y com distribuição normal assimétrica com média $\mu = \mu_{\phi_{kt}}$, variância $\sigma^2 = \sigma_{\phi_{kt}}^2 / t$ e constante de assimetria $\lambda = (\mu - x_{\phi_{kt}}) / t$, ou seja,

$$Y \sim SN(\mu_{\phi_{kt}}, \sigma_{\phi_{kt}}^2 / t, (\mu - x_{\phi_{kt}}) / t).$$

PASSO 3: Selecione um número aleatório y da distribuição Y e atualize o valor da componente ϕ_{kt} com o valor y^* , sendo y^* a parte inteira de y .

Com o método de mutação que foi apresentado, é possível observar que o novo valor da componente ϕ_{kt} é proveniente de uma distribuição normal assimétrica com média igual ao próprio valor atual da componente, porém com variabilidade e assimetria que decaem com o número de gerações t do AGHM.

A grande vantagem em utilizar a distribuição normal assimétrica ao invés da distribuição normal é diminuir a chance de que o AGHM, nas gerações iniciais, visite locais mais afastados que o atual valor da componente, justamente porque o grau de assimetria λ , nesse início, tende a ser maior. E à medida que as gerações de soluções vão sendo criadas esse raio de visitas vai diminuindo porque tanto o λ quanto σ^2 vão diminuindo, deixando a distribuição cada vez mais simétrica e cada vez mais concentrada em torno da componente ϕ_{kt} .

Esses dois aspectos de variabilidade e grau de assimetria diminuindo com o aumento de número de gerações torna o método de mutação muito atrativo e, de uma certa forma, com uma característica heurística, em que a própria técnica vai “aprendendo” a mutar os seus genes de uma forma cada vez menos variável, o que provoca a produção de descendentes cada vez mais adaptáveis.

5.3.7. RECOMBINAÇÃO

A recombinação (*crossover*) consiste em criar descendentes, de tal forma que algumas características dos parentes sejam trocadas entre si (MICHALEWICZ, 1994). A idéia é que se tenha a chance de criar indivíduos melhores e mesmo que se criem também piores, esses terão menor representação nas seleções seguintes. Além disso, para cada componente do vetor cromossomo, o operador *crossover* ocorrerá com uma probabilidade P_r .

Para a representação real, o operador recombinação será desenvolvido a partir de uma técnica semelhante ao que foi apresentado para o operador mutação, de modo a produzir indivíduos mais próximos daqueles mais aptos, no sentido de terem um melhor valor da função *fitness*. Nesse sentido, sabe-se que a *k*-ésima componente de dois indivíduos-pais foi selecionada, considere o seguinte procedimento para a realização do processo de recombinação:

PASSO 1: Sejam ϕ_{kt}^{p1} e ϕ_{kt}^{p2} os valores da *k*-ésima componente de dois indivíduos-pais com as respectivas funções de *fitness* g_t^{p1} e g_t^{p2} . O ajuste reflete a variância *a posteriori* de *p*, então o ideal é construir uma combinação linear das *k*-ésimas componentes dos vetores-pais, de modo que favoreça a transferência do material genético do pai que apresentar o menor valor para a função de ajuste. Essa combinação linear é dada por:

$$\mu_{\phi_k} = \frac{\left(\frac{1}{g_t^{p1}} \phi_{kt}^{p1} + \frac{1}{g_t^{p2}} \phi_{kt}^{p2} \right)}{\left(\frac{1}{g_t^{p1}} + \frac{1}{g_t^{p2}} \right)} \quad (5.17)$$

PASSO 2: Sejam LI_{ϕ_k} e LS_{ϕ_k} os limites da k -ésima componente dos indivíduos-pais na geração t , obter o desvio padrão (σ_{ϕ_k}) e o ponto médio do intervalo (x_{ϕ_k}), respectivamente, como segue:

$$\begin{aligned}\sigma_{\phi_k} &= \frac{LS_{\phi_k} - LI_{\phi_k}}{6} \\ x_{\phi_k} &= \frac{LI_{\phi_k} + LS_{\phi_k}}{2}\end{aligned}\tag{5.18}$$

PASSO 3: Defina uma variável aleatória Y com distribuição normal assimétrica com média $\mu = \mu_{\phi_k}$, variância $\sigma^2 = \sigma_{\phi_k}^2 / t$ e constante de assimetria $\lambda = (\mu - x_{\phi_k}) / t$, ou seja, $Y \sim SN(\mu_{\phi_k}, \sigma_{\phi_k}^2 / t, (\mu - x_{\phi_k}) / t)$.

PASSO 4: Selecione um número aleatório y da distribuição Y . Aquele indivíduo-pai que apresentou a menor função de *fitness* terá a sua k -ésima componente atualizada com o novo valor y^* , sendo y^* a parte inteira de y , e será o descendente dos dois vetores-pais que seguirá para as gerações subsequentes.

Novamente, nota-se que o método de recombinação sugerido acima apresenta traços fortes dos conceitos heurísticos, isto é, das ferramentas evolutivas que privilegiam os indivíduos mais adaptados ao meio. Isso garante uma convergência mais rápida do algoritmo genético híbrido multiobjetivo. Ainda assim, de ciclos em ciclos, é necessário uma busca local multiobjetivo pelos melhores indivíduos que ainda não foram examinados. Esse procedimento é abordado na seção seguinte.

5.3.8. BUSCA LOCAL MULTIOBJETIVO

No algoritmo genético AGHM, aplicou-se um procedimento de busca local para melhorar as soluções dominantes da população P'_t gerada pelos operadores genéticos. O procedimento da busca local multiobjetivo (BLM), descrito no Algoritmo 6, é executado a partir do conjunto S de soluções dominantes da população P'_t .

ALGORITMO 6: BUSCA LOCAL MULTIOBJETIVO (BLM)

Entrada: S (um conjunto de soluções dominantes).
 N_{max} (número máximo de caminhos a explorar).
 $Niter$ (número máximo de iterações)
Saída: D_{BL} (conjunto de soluções dominantes melhoradas).

PASSO 1: *Início*

Faça $t = 1$ (contador de iterações) e $D_{BL} = S$.

PASSO 2: *Redução do conjunto S*

Se $|S| > N_{max}$ então “ $|S|$ é o número de elementos do conjunto S ”
Reduza o conjunto S selecionando somente N_{max} soluções.
Seja LC o conjunto das soluções não selecionadas.

PASSO 3: *Construção de um conjunto de soluções vizinhas S'*

Faça $S' = \emptyset$ (conjunto das soluções vizinhas dominantes).

Para cada $x_k \in S$ faça

Gere a vizinhança de x_k : $N(x_k)$.

Para cada $y \in N(x_k)$ faça

Se $y \notin D_{BL}$ ou y não é dominado por D_{BL} então

Faça $S' =$ soluções dominantes de $(S' \cup \{y\})$.

PASSO 4: *Atualização do conjunto D_{BL} com as novas soluções vizinhas em S'*

$D_{BL} =$ soluções dominantes de $(D_{BL} \cup S')$.

PASSO 5: *Faça S' = soluções dominantes de (S' ∪ LC)*

Se $(S' \neq \emptyset$ e $t \leq Niter)$ então

Faça $S = S'$, $t = t + 1$ e volte ao passo 1.

Senão, pare.

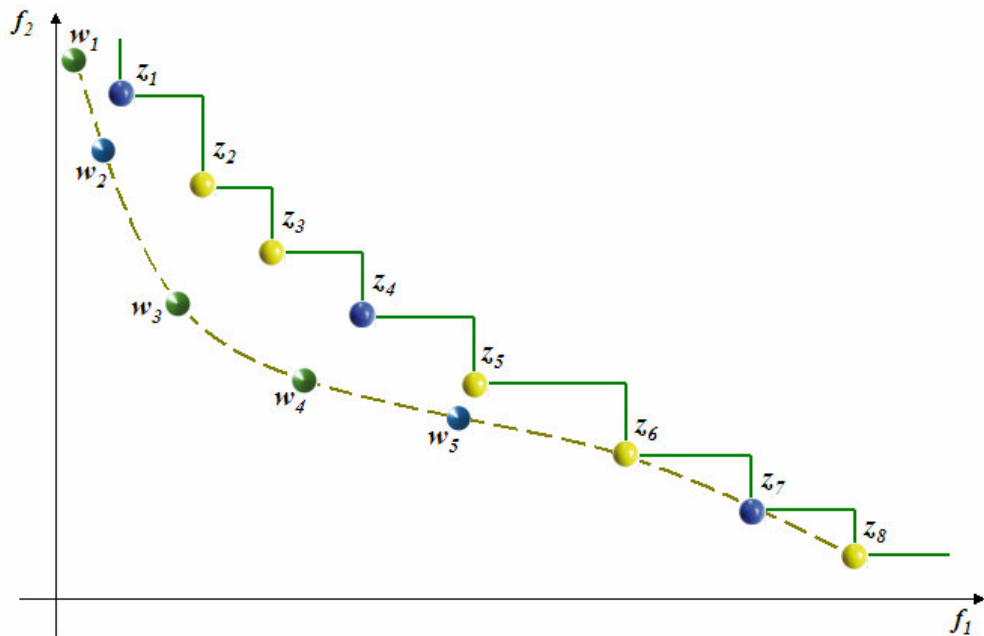


FIGURA 5.13: Ilustração gráfica do método de busca local multiobjetivo – BLM

A FIGURA 5.13 ilustra um exemplo de aplicação do procedimento de busca local multiobjetivo que será detalhado a seguir.

Assim, seja $S = \{x_1, x_2, \dots, x_8\}$ o conjunto de partida da busca local e suponha que $N_{max} = 3$ seja o número máximo de caminhos a serem explorados. A FIGURA 5.13 destaca a fronteira $f(S)$ formada pelos pontos $z_i = f(x_i)$, $i = 1, \dots, 8$. Inicialmente, o conjunto de soluções dominantes $D_{BL} = S$. Conforme o passo 2 do Algoritmo BLM, o conjunto S deve ser reduzido a três soluções. Suponha que x_1 , x_4 e x_7 sejam as três soluções selecionadas para serem exploradas pela busca local, ou seja, o conjunto inicial da busca local é $S = \{x_1, x_4, x_7\}$, e as soluções não selecionadas são armazenadas no conjunto $LC = \{x_2, x_3, x_5, x_6, x_8\}$. A partir do conjunto S constrói-se o conjunto S' com cinco soluções vizinhas, $S' = \{y_1, y_2, y_3, y_4, y_5\}$ (veja na FIGURA 5.14 os pontos imagens $w_i = f(y_i)$, $i = 1,$

..., 5). Note que, S' sempre conterá soluções não dominadas pelos conjuntos S e D_{BL} . O conjunto D_{BL} é atualizado fazendo com que esse receba as soluções dominantes de $(D_{BL} \cup S') = \{y_1, y_2, y_3, y_4, y_5, x_6, x_7, x_8\}$ (passo 4 do Algoritmo BLM).

No passo 5 do Algoritmo BLM, as soluções que não foram selecionadas para busca local podem ser consideradas na próxima iteração se essas não são dominadas por S' ; então, S' receberá as soluções dominantes de $(S' \cup LC) = \{y_1, y_2, y_3, y_4, y_5, x_6, x_8\}$. Como o conjunto inicial S foi melhorado pelo conjunto $S' \neq \emptyset$, uma vez que as soluções em $f(S')$ dominam as soluções em $f(S)$, o processo é reiniciado numa nova iteração fazendo $S = S'$. Desse modo, no passo 2 dessa nova iteração, os pontos selecionados para serem explorados podem ser y_2, y_5 e x_8 que correspondem aos pontos w_2, w_5 e z_8 .

Observe que a busca local BLM termina pela execução de um número de iterações ou quando não são encontradas novas soluções dominantes, isto é, o conjunto D_{BL} não é alterado. Se no passo 3 nenhuma solução é adicionada a S' o algoritmo tentará encontrar novas soluções dominantes a partir das soluções armazenadas em LC , finalizando quando $LC = \emptyset$.

A busca local BLM explora vários caminhos simultaneamente, e o conjunto final obtido por este método contém várias soluções dominantes que constituem uma aproximação da fronteira das soluções Pareto-ótimas, cuja qualidade depende da definição da vizinhança e do conjunto inicial de soluções dominantes.

Foi justamente com a intenção de melhorar a qualidade do algoritmo BLM que se decidiu utilizar uma rede neural. A tecnologia de redes neurais mimetiza o processo de resolução de problemas do cérebro humano. Tal como humanos aplicam o conhecimento adquirido de experiências passadas para novos problemas ou situações, uma rede neural utiliza exemplos

resolvidos previamente para construir um sistema de "neurônios" que toma novas decisões, faz classificações e prognósticos.

Em resumo, as redes neurais artificiais podem ser definidas como um sistema de processamento da informação que tem características de desempenho em comum com as redes neurais biológicas. Elas são capazes de solucionar problemas de inteligência artificial, construindo um sistema que tenha circuitos que simulem o cérebro humano, inclusive seu comportamento, ou seja, aprendendo, errando e fazendo descobertas. São mais que isso, são técnicas computacionais que apresentam um modelo inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

Os trabalhos sobre redes neurais se iniciaram na década de 40, com o neurofisiologista McCulloch e o matemático Pitts (McCULLOCH; PITSS, 1943). Eles estabeleceram uma analogia entre os processos de comunicação das células nervosas e de transmissão elétrica por impulsos, e propuseram a criação dos neurônios, os quais eram capazes de se organizarem em redes executando diversas tarefas. No entanto, o modelo que se tornou mais popular foi a rede *perceptron*, criada por Frank Rosenblatt (1958), baseando-se no aprendizado supervisionado por correção de erros. Mas foram as redes multiníveis, conhecidas como retropropagação (*back-propagation*), que transformaram as redes *perceptron* em sistemas generalistas, pois permitiam a correção de erros em redes contendo múltiplas camadas (RUMELHART *et al.*, 1986b).

A forma de arranjar os *perceptrons* em camadas é denominada *perceptron* multicamadas (PMC). O *perceptron* multicamadas foi concebido para resolver problemas mais complexos, os quais não poderiam ser resolvidos pelo modelo de neurônio básico. Para isso são necessárias mais conexões, os quais só existem em uma rede de *perceptrons* dispostos em camadas. Tipicamente, a rede PMC consiste de um conjunto de unidades sensoriais (nós de fonte) que constituem a camada

de entrada, uma ou mais camadas ocultas de nós computacionais e uma camada de saída de nós computacionais. O sinal de entrada se propaga para frente através da rede, camada por camada (HAYKIN, 2001). Usualmente as camadas de um PMC são classificadas em três grupos

- **Camada de Entrada:** onde os padrões são apresentados à rede;
- **Camadas Intermediárias ou Ocultas:** onde é feita a maior parte do processamento, através das conexões ponderadas; podem ser consideradas como extratoras de características;
- **Camada de Saída:** onde o resultado final é concluído e apresentado

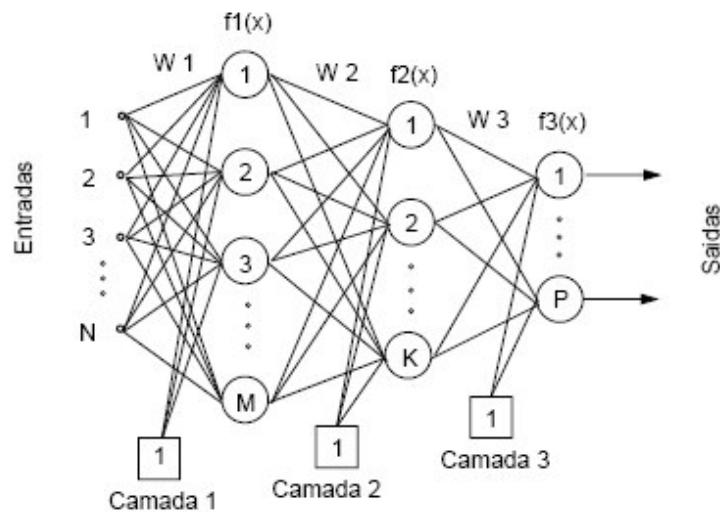


FIGURA 5.14: Modelo de uma rede neural do tipo PMC

Portanto, para tentar reduzir o tempo de processamento da busca local multiobjetivo decidiu-se utilizar uma rede neural tipo *perceptron* com 2 camadas escondidas para elencar ou selecionar as soluções dominantes iniciais mais aptas, de modo a aumentar a eficiência da busca local

(ROSENBLATT, 1958). A idéia é tentar escolher, por meio da rede neural, o melhor ponto de partida para o método BLM, isto é, o melhor conjunto inicial de soluções dominantes. Essa técnica, entre outros benefícios, tende a reduzir significativamente o tempo total para a obtenção do vetor ótimo de soluções.

Ainda assim, para evitar que a busca local consuma a maior parte do tempo computacional do algoritmo genético, sugere-se acionar a busca local somente a cada β iterações do algoritmo genético. Como a busca local explora vários caminhos em paralelo, é necessário definir, de forma apropriada, os parâmetros N_{max} (número máximo de caminhos a explorar) e $Niter$ (número máximo de iterações da busca local).

Ao finalizar a busca local BLM, deve-se atualizar a população P'_t com o conjunto D_{BL} de soluções obtido por BLM. Seja $|D_{BL}|$ o número de soluções contidas no conjunto D_{BL} . Então, as soluções de D_{BL} devem ser inseridas na população e $|D_{BL}|$ soluções da população são removidas. Como as soluções de D_{BL} foram obtidas a partir de S , uma alternativa é substituir S por D_{BL} . No entanto, existem três casos para os tamanhos destes conjuntos a serem analisados: $|D_{BL}| = |S|$, $|D_{BL}| > |S|$ ou $|D_{BL}| < |S|$. Apresenta-se a seguir um algoritmo para atualizar a população P'_t com o conjunto D_{BL} obtido pela busca local multiobjetivo.

ALGORITMO 7: ATUALIZAÇÃO DA POPULAÇÃO COMO O CONJUNTO DOMINANTE D_{BL} (ADBL)

Sejam $S = \{x_i : i = 1, \dots, |S|\} \subset P'_t$ o conjunto de soluções dominantes da população P'_t antes da busca local e $DBL = \{y_i : i = 1, \dots, |DBL|\}$ o conjunto de soluções dominantes obtido pela busca local multiobjetivo.

Entrada: S (um conjunto de soluções dominantes).

D_{BL} (conjunto de soluções dominantes melhoradas).

Saída: S (um conjunto de soluções dominantes) ainda melhor.

PASSO 1: Se $|DBL| = |S|$ e $DBL \neq S$ então faça $S = DBL$.

Senão

Se $|DBL| > |S|$, então

Para $i = 1$ até $|S|$ faça $x_i = y_i$ (insira y_i em S e descarte x_i).

Para $i = |S| + 1$ até $|DBL|$ faça

Selecione aleatoriamente uma solução x' de $(P'_t - S)$ e faça $x' = y_i$.

Faça $S = S \cup \{x'\}$ (para que x' não seja selecionada novamente).

Senão

Ordene aleatoriamente as soluções em S .

Para $i = 1$ até $|DBL|$ faça $x_i = y_i$ (note que esta substituição é aleatória).

FIGURA 5.15: Algoritmo para atualizar a população P'_t por meio do algoritmo de busca local multiobjetivo

A FIGURA 5.16 demonstra um exemplo de atualização da população P'_t em que $f(S)$ e $f(DBL)$ são, respectivamente, as imagens dos conjuntos S e DBL no espaço objetivo e $z_i = f(x_i)$, $\forall i$. No caso em que $|DBL| > |S|$ (FIGURA 5.16a), todas as soluções de S são substituídas por soluções de DBL , e as soluções restantes de DBL substituem soluções selecionadas aleatoriamente de $(P'_t - S)$. Por outro lado, na situação em que $|DBL| < |S|$ (FIGURA 5.16b), $|DBL|$ soluções de S são substituídas pelas soluções de DBL , e as outras soluções de S , permanecem intactas na população.

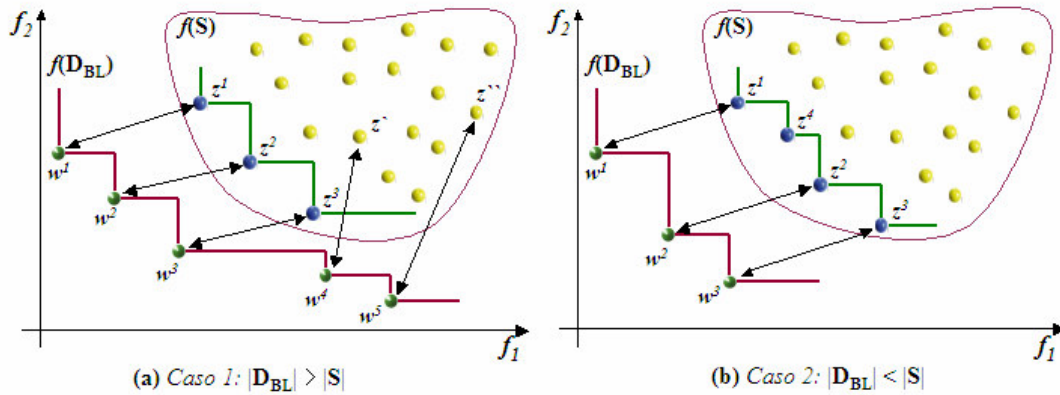


FIGURA 5.16: Atualização da população P'_t a partir do conjunto DBL obtido pela busca local

Vale ressaltar que, todos os pontos z^i explorados pela busca local são armazenados numa lista W para evitar que a busca local explore novamente esses pontos. Além disso, sabendo que a lista W pode crescer rapidamente, então o algoritmo armazena somente os pontos dominantes daqueles já explorados. Portanto, se numa dada iteração um ponto $z \in f(S)$ (z ponto dominante da população) pertence também à lista L , então z deve ser descartado do conjunto $f(S)$, pois ele já foi explorado em alguma iteração anterior. Note que S é o conjunto de soluções de partida da busca local.

5.3.9. PSEUDOCÓDIGO DO ALGORITMO GENÉTICO HÍBRIDO MULTIOBJETIVO

O pseudocódigo mostrado a seguir reúne todos os conceitos vistos nessa seção e apresenta todas as suas fases no processo de otimização multiobjetivo.

ALGORITMO 8: ALGORITMO GENÉTICO HÍBRIDO MULTIOBJETIVO

Entrada: N (tamanho da população)
 p_R, p_M (probabilidade de recombinação e mutação)
 $MaxE$ (número máximo de soluções de elite)
 N_{max} (número máximo de caminhos a serem exploradas pela busca local)
 N_{iter} (número máximo de iterações da busca local)
 β (período de ativação da busca local. Sugestão: $\beta > 3$)

Saída: D (conjunto de soluções dominantes)

PASSO 1: Início

- Gerar uma população inicial P_1 de N soluções (as soluções em P_1 podem ser geradas aleatoriamente ou usando alguma heurística construtiva).
- Para cada solução em P_1 , calcule os valores dos r objetivos.
- Inicializar o conjunto de soluções dominantes: $D_1 = \emptyset$.
- Inicializar o conjunto de soluções de elite: $E_0 = \emptyset$.
- Seja W uma lista contendo as soluções já exploradas pela busca local. Faça $W = \emptyset$ e $t = 1$.

PASSO 2: Classificação da População

Classificar a população $P_t \cup E_{t-1}$ de acordo com a relação de dominância das soluções obtendo uma lista de K conjuntos dominantes F_1, F_2, \dots, F_K .

PASSO 3: Atualização da População

Faça $D_t =$ soluções dominantes de $(D_t \cup F_1)$.

PASSO 4: Cálculo do Fitness

Para cada $x \in F_i$ ($i = 1, 2, \dots, K$) calcule $fitness(x)$.

PASSO 5: Seleção

Seja $SP_t = \emptyset$. Usando um método de seleção, escolha N soluções de $P_t \cup E_{t-1}$ e armazene-os em SP_t .

PASSO 6: *Atualização do conjunto de soluções de elite*

Se $|D_t| > MaxE$, então faça $N_{elite} = MaxE$, senão faça $N_{elite} = |D_t|$.
Selecionar aleatoriamente N_{elite} soluções de D_t e armazene-as em E_t .

PASSO 7: *Recombinação e Mutação*

Seja $P'_t = \emptyset$ uma nova população.

Repita $N/2$ vezes

- Escolha duas soluções (pais) $x_1, x_2 \in SP_t$ e faça $SP_t = SP_t - \{x_1, x_2\}$. Recombine x_1 e x_2 aplicando um operador de recombinação.
- Adicione a P'_t as novas soluções (filhos) y_1 e y_2 com probabilidade p_R . Caso contrário, adicione x_1 e x_2 a P'_t .

Para cada solução $x \in P'_t$ faça:

 Aplique a x um operador de mutação com probabilidade p_M .

PASSO 8: *Avaliação*

Para cada solução em P'_t , calcule os valores dos r objetivos.

PASSO 9: *Looping*

Se $t = 1$ ou t é um múltiplo de β , então vá para o passo 10.

Senão, faça $t = t + 1$, $P_t = P'_{t-1}$ e vá para o passo 12.

PASSO 10: *Determinação das soluções dominantes de P'_t a serem exploradas pela busca local multiobjetivo*

S = soluções dominantes de P'_t .

Para cada solução $x \in S$ faça

 Se $x \in W$ (i.e., x já foi explorado pela busca local) então remova x de S .

Atualize W com os pontos a serem explorados: $W =$ soluções dominantes de $(W \cup S)$ (note que, W armazena somente os pontos dominantes de todos os pontos já explorados pela busca local).

PASSO 11: *Busca Local Multiobjetivo*

Execute a busca local multiobjetivo (Algoritmo 1) a partir do conjunto S (utilize os parâmetros N_{max} e N_{iter}). Seja D_{BL} o conjunto de soluções dominantes obtidos ao final da busca local.

Atualize a população P'_t com D_{BL} (Algoritmo 8)

PASSO 12: *Teste de Parada*

Se a condição de parada é satisfeita, então faça $D =$ soluções dominantes de $(D_t \cup P_t)$ e pare.

Senão, faça $t = t + 1$, $P_t = P'_{t-1}$ e retorne ao passo 2.

No algoritmo genético AGHM, para selecionar as soluções reprodutoras (pais), pode ser usado o método da roleta (*roulette wheel*), como descrito em Goldberg (1989), que consiste no sorteio de soluções baseado nos valores dos *fitness*, isto é, as soluções com os menores valores de *fitness* possuem maiores chances de serem escolhidas. Cabe destacar que na implementação da metaheurística AGHM o *fitness* de uma solução é obtido a partir do valor normalizado da distância de aglomeração, como está sendo proposto aqui.

Sobretudo, é a combinação de diversas técnicas, tais como algoritmos genéticos com busca local multiobjetivo e redes neurais, que constitui a contribuição mais importante da metaheurística AGHM ora discutida. Além disso, alguns testes foram feitos com a técnica AGHM e verificou-se que a mesma apresenta uma taxa de convergência, em termos de tempo, da ordem de $O(N^2)$. Vale lembrar ainda que, na literatura de otimização multiobjetivo, a maioria dos algoritmos genéticos propostos não utiliza estratégias de busca local multiobjetivo. A FIGURA 5.17 ilustra o processo geral de otimização multiobjetivo discutido nesse texto.

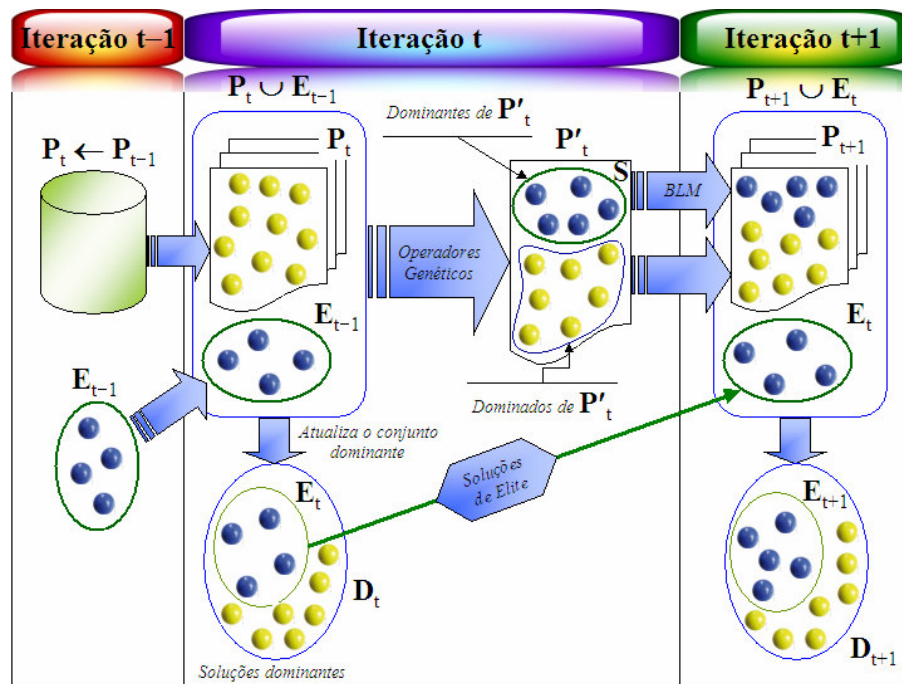


FIGURA 5.17: Visão geral da implementação do algoritmo genético híbrido multiobjetivo

5.4. APLICAÇÃO A UM EXEMPLO NUMÉRICO

Para ilustrar o funcionamento do algoritmo genético híbrido multiobjetivo proposto na seção 5.3, bem como o seu desempenho, serão utilizados, mais uma vez, os dados analisados no artigo de Fujisawa e Izume (2000), cujo objetivo era estimar o percentual verdadeiro de alelos de quatro antígenos (M, N, S, s) para as cidades de Hiroshima e Nagasaki.

Em especial, para a aplicação do AGHM serão considerados os dados referentes ao antígeno M da cidade de Nagasaki (TABELA 5.1) com três classificações repetidas, no qual foram realizadas 156 investigações, sendo que 26 não apresentaram respostas positivas entre os três testes realizados; dois apresentaram exatamente uma resposta positiva; um paciente obteve dois resultados positivos e 127 exames forneceu três resultados positivos.

Tabela 5.1: Frequência Observada de Respostas Positivas (k) entre os Testes Serológicos Repetidos (m) para o Antígeno M da cidade de Nagasaki

$k(m=3)$ e $n=156$			
0	1	2	3
26	2	1	127

Desse modo, adaptando o experimento às condições de entrada do algoritmo genético híbrido multiobjetivo, considerou-se uma Matriz-Trabalho de 156, que é o total de pacientes avaliados, e uma distribuição *a priori* Beta (1,2) para os erros de classificação. Para essa aplicação, formulou-se a seguinte questão: Para a cidade de Nagasaki, quantos pacientes seriam necessários examinar (n),

quantos testes serológicos deveriam ser realizados (m) em cada paciente e qual seria o melhor critério (a) para decidir se o paciente apresentou ou não uma resposta positiva, de modo que a estimativa sobre a proporção de alelos contendo o antígeno estudado apresentasse, ao mesmo tempo, uma variabilidade de, no máximo, 0,001 e com um intervalo de credibilidade de 95%, cujo seu comprimento não ultrapassasse 0,080? Vale ressaltar que os valores fixados tanto para a variabilidade máxima quanto para o comprimento máximo do intervalo de credibilidade são meramente ilustrados, objetivando apenas a utilização do algoritmo genético proposto neste capítulo. Numa situação real, é comum que tais parâmetros sejam definidos por uma equipe técnica conhecedora do problema que está sendo investigado.

A questão não é trivial porque o vetor ótimo de soluções (m, n, a) tem de, simultaneamente, gerar uma variância *a posteriori* para o parâmetro de interesse e um comprimento de intervalo de $100(1-\alpha)\%$ credibilidade inferiores aos respectivos valores pré-estabelecidos.

A TABELA 5.2 mostra os valores ótimos para o tamanho da amostra, o número de classificações repetidas e o critério para a classificação final de cada paciente que atenderam às especificações de manter tanto uma variância máxima de 0,001 para a distribuição *a posteriori* da proporção de alelos com o antígeno M da cidade de Nagasaki quanto uma amplitude máxima de 0,080 para um intervalo de credibilidade de 95%. O tempo total de processamento se aproximou de 10 horas.

Para esse antígeno da cidade de Nagasaki, Fujisawa e Izume (2000) encontraram uma estimativa de 0,820 para o verdadeiro percentual de alelos com limites de 0,757 e 0,872 para um intervalo de credibilidade de 95%. Enquanto que ao utilizar o AGHM, identificaram-se três resultados ótimos para o vetor (n, m, a) que atenderam às exigências iniciais.

Tabela 5.2: Valores ótimos do tamanho da amostra, o número de classificações repetidas e o critério para a classificação final de cada paciente

n	m	a	$\hat{\sigma}_p^2$ máxima	Amplitude Máxima do Intervalo MDP (95%)
39	4	1	0,0003755861	0,0760024053
52	3	1	0,0003400962	0,07400053986
52	3	2	0,0003177349	0,0718632388

Além disso, nota-se que, para manter a Matriz-Trabalho fixa em um esforço computacional de 156 avaliações, a dupla experimental (n, m) que parece mais adequada para se trabalhar é $n=39$, $m=4$ e $a=1$ como critério de classificação final para identificação do antígeno M no indivíduo.

Essa solução, no pior cenário possível, forneceu um intervalo de credibilidade de 95%, cuja amplitude é 34% menor em relação à amplitude do intervalo de credibilidade de 95% conseguido por Fujisawa e Izume (2000). Observou-se, também, que o tamanho amostral obtido foi 4 vezes menor com a aplicação de quatro testes serológicos. Cabe, entretanto ressaltar que Fujisawa e Izume (2000) utilizaram uma abordagem estatística clássica e não bayesiana.

Evidentemente, está se referindo a um estudo clínico prospectivo, isto é, se fosse necessário planejar um novo estudo no futuro como o que foi realizado pelos autores, bastaria investigar 39 pacientes, aplicando-os quatro testes serológicos com o critério de classificação final $a=1$ para se atingir, ao mesmo tempo, uma variância *a posteriori* para a proporção de alelos contendo o antígeno M de, no máximo, 0,001 e com um intervalo de credibilidade de 95% de amplitude máxima 0,080. Além do mais, esses resultados obtidos pela metodologia AGHM apresentaram muita consistência e robustez, uma vez que refletem a configuração do pior cenário possível em relação às duas medidas

avaliadas, ou seja, no pior cenário possível, a variância e a amplitude do intervalo de credibilidade para estimar tal proporção seriam, no máximo, àqueles valores especificados.

Por outro lado, caso seja de interesse não restringir o processo de otimização por uma determinada Matriz-Trabalho, é possível relaxar essa condição. Num processo mais geral, talvez o critério de controle não fosse mais a Matriz-Trabalho, mas sim um tamanho de amostra máximo e um número máximo de classificações repetidas que refletissem com mais exatidão uma situação real. Essa idéia será tratada na seção 5.5.

5.5. DELINEAMENTO DO PROCESSO DE SIMULAÇÃO

O processo de simulação está fundamentado no algoritmo genético híbrido multiojetivo proposto na seção 5.3. Os parâmetros de entrada são apresentados e discutidos a seguir:

CRITÉRIOS DO PROCESSO

1. **Tamanho da Matriz-Trabalho (R):** Designa todo o esforço computacional utilizado para obter o veto ótimo (n, m, a) . Lembre-se que uma Matriz-Trabalho é obtida como $n \times m$;
2. **n_{max} e m_{max} :** Designam o tamanho máximo amostral e o número máximo de classificações repetidas que devem ser considerados no processo de otimização. Esses parâmetros conjuntamente e o tamanho da Matriz-Trabalho (R) são mutuamente excludentes, isto é, ou é fornecido o valor de R ou são fornecidos os valores de n_{max} e m_{max} ;
3. **σ_{max}^2 :** Variância máxima permitida da distribuição a posteriori de p ;
4. **α :** Significância para a construção do intervalo de credibilidade;
5. **w_{max} :** Amplitude máxima do intervalo MDP para uma credibilidade de $100(1-\alpha)\%$;

CRITÉRIOS DO ALGORITMO

6. P_m e P_r : São as probabilidades de mutação e de recombinação do algoritmo genético;
7. IT_{max} : Número máximo de iterações (gerações) utilizado pelo AGHM;
8. **Critério de parada:** O AGHM parará quando, pelo menos, um dos critérios for observado:
 - a. IT_{max} foi atingido;
 - b. Todos os elementos da Matriz-Trabalho de tamanho R pesquisados;
 - c. Os valores de n_{max} e m_{max} foram atingidos;

Nota-se que ao se definir a Matriz-Trabalho como tendo tamanho R , uma vez que $R=n \times m$, automaticamente o AGHM definirá o seu espaço de busca multiobjetivo como sendo todos os valores de n que são divisores de R , de tal forma que $m=R/n$.

Em algumas situações que não exigem uma busca tão minuciosa da solução ótima, essa estratégia funciona bem e de forma muito ágil. Por outro lado, às vezes, a melhor solução será utilizada como ferramenta de tomada de decisão e, então, uma resposta ótima mais criteriosa deve ser apresentada. Nesse caso, é mais interessante definir o tamanho amostral máximo (n_{max}) e o número máximo de classificações repetidas (m_{max}), e, portanto, o espaço de busca do algoritmo genético híbrido multiobjetivo será todos os valores possíveis, e inteiros, envolvendo o produto $i \times j$, sendo $i=1, 2, \dots, n_{max}$ e $j=1, 2, \dots, m_{max}$. Para cada combinação $i \times j$, o AGHM consultará, de forma estocástica, alguns ou até todos os a 's, tais que $1 \leq a \leq j$. Essa estratégia deixa o algoritmo proposto mais lento, porém com uma precisão maior, pois potencializa a obtenção de soluções mais próximas da região de pontos Pareto-ótimo. Será utilizado o pseudocódigo mostrado no Algoritmo 8 considerando dois estudos de simulação.

Tabela 5.3: Definição dos parâmetros de entrada do AGHM para dois estudos propostos

ESTUDO 1:	ESTUDO 2:
1. $R=600$	1. $n_{max}=600$ e $m_{max}=7$
2. $P_m=0,01$ e $P_r=0,80$	2. $P_m=0,01$ e $P_r=0,80$
3. $\sigma^2_{max}=0,0003$	3. $\sigma^2_{max}=0,0003$
4. $\alpha=0,05$	4. $\alpha=0,05$
5. $w_{max}=0,01$	5. $w_{max}=0,01$
6. $IT_{max}=10000$	6. $IT_{max}=10000$

O primeiro estudo privilegia o tempo de processamento das soluções factíveis a serem encontradas, já que o algoritmo se concentrará somente nos 24 pares (n, m) de divisores da Matriz-Trabalho com tamanho 600 indivíduos. Para cada um desses pares, pelo menos, um dos critérios possíveis (a) para a classificação final de um determinado item será testado como componente do vetor ótimo multiobjetivo, ou seja, para cada n , o número de classificações repetidas será calculado como $m=R/n$ e, conseqüentemente, com certa freqüência, alguns critérios a , limitados a m , serão testados. Ainda assim, o esforço computacional dessa estrutura de simulação é bem reduzido e o resultado será um conjunto de soluções que poderá não atender às exigências operacionais de um determinado sistema produtivo.

Já o estudo 2 que não apresenta a Matriz-Trabalho, é baseada nos valores máximos de n (n_{max}) e de m (m_{max}). Isso quer dizer que $\forall n_{max} \leq 100$ e $m_{max} \leq 15$, o algoritmo genético poderá investigar até 1500 pontos ($n_{max} \times m_{max}$) do espaço de soluções, sendo que, para cada um desses, ao menos, um critério para a classificação final de um determinado item será avaliado. O total de pares (i, j) testados dependerá dos elementos selecionados pelo algoritmo de busca local multiobjetivo (Algoritmo 6) e dos pontos pesquisados devido aos operadores de mutação e recombinação.

Portanto, por visitar mais soluções, o estudo 2 apresentará, no mínimo, um conjunto mais amplo de soluções. Para se ter uma idéia, para $n \leq 100$, as soluções encontradas pelo estudo 1, com uma grande chance, deverão estar contidas no conjunto de soluções geradas pelo estudo 2, porque o algoritmo de busca local multiobjetivo investiga todas as soluções vizinhas ao conjunto de elite de soluções factíveis obtidas em cada geração.

Sobre as questões relacionadas ao cálculo do tamanho de amostra com dados binomiais envolvendo erros e classificações repetidas existem poucas obras. Apesar disso, um grupo de pesquisadores canadenses tem desenvolvido alguns trabalhos. Rahme *et al.* (2000), por exemplo, propuseram uma metodologia, embora mais simples, semelhante ao estudo 2. Os autores não utilizaram algoritmos evolutivos e o problema se resumia em encontrar o menor tamanho amostral, tal que o intervalo MDP de maior amplitude possível tivesse cobertura de, no mínimo, $100(1-\alpha)\%$, ou seja, era um problema de otimização uniobjetivo.

Além disso, o algoritmo de busca foi implementado de forma incremental sobre n . Então, o processo se iniciava com $n^* = 1$ e o critério de parada era testado, ou seja, eram gerados todos os tamanhos de intervalos de credibilidade e a área da distribuição *a posteriori* baseada no maior desses intervalos (w_{max}^*) era comparada com a cobertura mínima de $1-\alpha$. No caso dessa área apresentar um valor inferior ao da probabilidade de cobertura estabelecida, o tamanho amostral era incrementado para $n^* = 2$ e todo o processo era repetido. A solução ótima será o menor n^* que atender o critério de parada do algoritmo.

Sob esse ponto de vista, por não ser incremental, o estudo 2 se diferencia do procedimento adotado por Rahme *et al.* (2000), pois o algoritmo apresentado aqui só será finalizado quando o conjunto de todos os pares (i, j) selecionados pela busca local multiobjetivo não mais puder ser melhorado, ou seja, quando não mais existir uma combinação de n , m e a que forneça uma variância

máxima *a posteriori* e um comprimento máximo de intervalo MDP com probabilidade $1-\alpha$ inferiores àqueles valores especificados previamente.

Esse último fato potencializa o leque de possibilidades para o tomador de decisão, uma vez que não necessariamente o menor tamanho amostral é a melhor solução para um determinado problema. Isto se deve ao fato que devemos considerar ainda o número de classificações repetidas m , bem como o critério a para classificação final do item e o risco (variância) total envolvido no processo.

Evidentemente, os dois estudos tratados aqui produzirão vetores ótimos diferentes, porque estão fundamentados em espaços de busca distintos e porque o algoritmo genético proposto é um processo estocástico de otimização. Ainda assim, será interessante comparar os valores ótimos para n , m e a obtidos por cada procedimento.

Análogo ao que foi descrito no Capítulo 4, o trabalho intensivo e computacional dispensado para a geração dos valores ótimos de n , m e a consumiu cerca de três dias para o estudo 1 e de sete dias aproximadamente para o estudo 2 e foi produzido na mesma máquina e plataforma já descritas. Tal fato revela um ponto positivo para a segunda abordagem, já que, mesmo tendo uma carga de cálculo maior, conseguiu analisar um número maior de possibilidades gastando-se um pouco mais que o dobro daquele registrado pelo estudo 1.

5.6. RESULTADOS NUMÉRICOS

Para a obtenção dos resultados, o algoritmo genético híbrido multiobjetivo proposto foi executado, para ambas as estratégias, com variabilidade *a posteriori* máxima da distribuição de p permitida de 0,00030 e com um intervalo de credibilidade MDP com amplitude máxima de 0,010 para um nível de credibilidade de 95%. Além disso, para ambas os estudos, foram utilizadas uma distribuição *a priori* Beta(1,2) para os erros de classificação.

A estrutura de otimização apresentada na equação 5.11 consiste em obter os menores valores de n e m , de modo que tanto a variância *a posteriori* máxima (σ_{max}^2) quanto o tamanho do intervalo de credibilidade máximo (w_{max}) estejam abaixo dos valores de referência. Dado que ficou decidido antes do experimento que ambos σ_{max}^2 e w_{max} representam valores aceitáveis, o objetivo do algoritmo é maximizar as funções $f_1(x)$ e $f_2(x)$ da equação 5.11. Em relação à economia, quanto menores forem n e m , mais viável será a solução.

Os resultados dos dois estudos estão descritos na TABELA 5.4, sendo que as soluções 5, 10, 11, 20 e 30 foram obtidas pelo estudo 1. Considerando então apenas essas soluções geradas pelo estudo 1, o menor tamanho amostral que atende às especificações é 150. Numa primeira análise, esse tamanho amostral condiz completamente com os preceitos teóricos de otimização já explicados, uma vez que, sob a ótica da economia, 150 é o menor tamanho amostral, cuja variância *a posteriori* de p e a amplitude do intervalo MDP, no pior cenário possível, são inferiores a 0,00030 e a 0,010, respectivamente.

Notou-se também que a sensibilidade das soluções factíveis às quantidades estatísticas calculadas *a posteriori* pode vir de várias outras fontes. Por exemplo, as soluções 2 e 3 retratam dois vetores ótimos muito próximos, diferenciando-os apenas pelo critério a . Mas a mudança de um

critério de classificação final de $a=1$ para $a=2$, considerando que foram feitas $m=3$ classificações repetidas, fará uma enorme diferença, pois reduzirá a variabilidade *a posteriori* em mais de 7% e o comprimento do intervalo em pouco mais de 12%.

Continuando ainda sob o ponto de vista do tomador de decisão ou do estrategista de mercado, note que nem sempre o vetor ótimo global (solução 1) será a solução adotada, pois existem outros fatores técnicos, econômicos e produtivos que, conjuntamente, agregam peculiaridades às diversas soluções encontradas que poderão torná-las mais atraentes ou inviáveis nos processos de manufatura. Assim, é possível optar por utilizar a solução 3, ao invés da solução 5, porque para um determinado sistema produtivo essa solução se mostrou mais promissora economicamente, por exemplo.

O mais importante é observar que, apesar de um conjunto limitado de soluções candidatas ter sido apresentado, a metaheurística híbrida aqui proposta fornece uma lista de soluções provavelmente boas, possibilitando ao tomador de decisão considerar sobre esses vetores ótimos candidatos outras questões subjetivas e pertinentes a cada área de conhecimento, de modo a permitir a configuração de vários cenários de mercado e de produção.

De imediato, esse fato já traz uma evolução para os problemas envolvendo o cálculo de tamanho da amostra em estudos práticos que necessitam dessa informação, uma vez que não somente uma solução é imposta, mas sim uma lista de várias soluções plausíveis não dominadas, isto é, com o mesmo grau de importância.

Por outro lado, o conjunto de soluções factíveis para o vetor (n, m, a) gerado pelo estudo 2 destaca-se, não só pela quantidade, mas também pela diversidade de resultados, conforme observa-se na TABELA 5.4. Vale lembrar, ainda, que para o estudo 2, não há a restrição que o trabalho computacional tem de ser 600, isto é, que $m \times n = 600$.

Por isso, em comparação com as soluções obtidas pelo estudo 1, comprova-se que o tamanho amostral ótimo poderia ser 128 (solução 1 da TABELA 5.4), ao invés de 150. Outras soluções entre 128 e 150 poderiam ser preferidas e escolhidas, mas não representariam necessariamente o pior cenário de estimação.

Tabela 5.4: Valores ótimos de n , m , e a para os parâmetros de entrada dos estudos 1 e 2, considerando uma variância *a posteriori* máxima de 0,00030 e uma amplitude máxima de 0,010 para um intervalo MDP de 95% de credibilidade.

Solução	n	m	a	$\hat{\sigma}_{p(máx)}^2$	$w_{máx}^*$	Solução	n	m	a	$\hat{\sigma}_{p(máx)}^2$	$w_{máx}^*$
1	128	4	1	0,0002999	0,0099502	16	279	5	2	0,0002429	0,0085098
2	138	7	2	0,0002894	0,0098298	17	286	6	2	0,0002419	0,0085027
3	143	7	1	0,0002841	0,0097298	18	291	5	4	0,0002415	0,0084881
4	149	5	3	0,0002838	0,0093619	19	300	2	1	0,0002413	0,0084859
5	150	4	2	0,0002810	0,0089810	20	300	4	2	0,0002410	0,0084825
6	159	6	5	0,0002799	0,0088961	21	346	4	1	0,0002197	0,0082452
7	163	7	2	0,0002767	0,0088029	22	376	5	2	0,0001839	0,0082443
8	189	5	1	0,0002748	0,0087893	23	398	7	6	0,0001804	0,0081434
9	192	5	2	0,0002725	0,0087380	24	417	4	1	0,0001769	0,0081042
10	200	3	1	0,0002710	0,0087018	25	469	4	3	0,0001645	0,0080408
11	200	3	2	0,0002517	0,0085940	26	511	3	1	0,0001552	0,0080260
12	203	5	4	0,0002504	0,0085930	27	553	7	2	0,0001471	0,0080181
13	228	6	2	0,0002475	0,0085422	28	572	5	4	0,0001160	0,0079842
14	243	7	5	0,0002469	0,0085259	29	596	4	2	0,0001044	0,0079629
15	255	7	5	0,0002433	0,0085149	30	600	1	1	0,0000967	0,0079389

Ainda assim, observa-se quantas soluções existem, além daquelas apresentadas pelo estudo 1, para variados tamanhos de amostra, dando assim, condições melhores para o tomador de decisão. Os cinco primeiros resultados gerados pelo estudo 2 formaram as cinco melhores soluções dentre todas as soluções obtidas utilizando ambos os estudos. Isto mostra que um algoritmo mais minucioso trará

resultados mais precisos porque consegue ser mais específico na obtenção das soluções factíveis, muito embora, o preço computacional seja, irremediavelmente, maior.

De um modo geral, esse tópico trouxe uma evolução para o campo de determinação de tamanho de amostras em estudos de dados dicotômicos com erros de diagnóstico e classificações repetidas, uma vez que foram usados algoritmos genéticos híbridos em problemas de determinação de tamanhos de amostra. Além disso, o problema com abordagem multiobjetivo foi estudado, ou seja, encontrar soluções provavelmente boas para um vetor de variáveis de decisão, de modo que, ao mesmo tempo, atendessem duas restrições de ordem técnica. O procedimento poderia ser generalizado para outros objetivos além da variância e amplitude do intervalo de credibilidade. Finalmente, ainda há toda capacidade e flexibilidade que os algoritmos genéticos híbridos têm de se adaptarem às mais diversas e complicadas situações práticas, restando ao tomador de decisão escolher um ponto de equilíbrio entre um algoritmo rápido e um algoritmo preciso.

Capítulo 6

CONCLUSÕES E EXTENSÕES

Este capítulo discorre sobre os principais resultados obtidos na tese e apresenta sugestões de projetos para pesquisas futuras.

6.1. CONCLUSÕES

Em muitas situações práticas, dados binários são utilizados para respaldarem tomadas de decisão. Esses dados podem estar sujeitos a erros de julgamento e não há um sistema classificador que seja infalível e mesmo que existisse ele poderia ser muito caro. O efeito dos erros de classificação tem sido estudado utilizando vários modelos e em diferentes circunstâncias. Em geral, a conclusão é que os erros de julgamento resultam em estimativas viciadas, como por exemplo, a probabilidade de conformidade em investigações por atributos.

Uma maneira de amenizar a influência desses erros é promover uma seqüência de m classificações repetidas para cada item avaliado, ao invés, talvez, de aumentar o tamanho da amostra. Bross (1954) foi pioneiro no estudo da influência dos erros de classificação, enquanto que Lindsey (1985) propôs um modelo baseado numa função de verossimilhança Completa (FVC). Evans *et al.* (1996) foram uns dos primeiros autores a utilizar análise bayesiana aplicada a dados dicotômicos sujeitos a erros de classificação com classificações repetidas. Mais recentemente, Pires (2006) e

Moura (2007) apresentaram algumas soluções para estudar a distribuição *a posteriori* da proporção de itens conformes (p) na presença de erros de diagnóstico.

E nesse contexto foi proposta uma metodologia para estimar o parâmetro p em dados binários sujeitos a erros e classificações repetidas. A técnica se utilizou de uma função de verossimilhança ordenada (FVO) dos dados amostrais e foi comparada com o modelo FVC, o que envolveu os cálculos das estimativas modais e dos desvios padrão das distribuições *a posteriori* de p . Uma análise de sensibilidade envolvendo três tamanhos de amostra, três tamanhos de classificações repetidas, três valores de proporção, três valores de erros de classificação e duas distribuições *a priori* para os erros, totalizando a investigação de 162 cenários.

Os resultados foram comparados e concluiu-se que o modelo proposto nesta tese é superior considerando o critério da variância *a posteriori* de p e do fator de Bayes. Sob essa última medida, a superioridade do modelo FVO em relação ao modelo FVC foi surpreendente, independentemente do recorte amostral, por n , m , p ou distribuição *a priori*, que foi efetuado. Houve casos em que o modelo proposto apresentou fator de Bayes próximo de 25 caracterizando uma grande evidência de superioridade do modelo proposto em relação à proposição de Evans *et al.* (1996).

Além disso, também para qualquer configuração experimental, constatou-se que as razões das amplitudes dos intervalos de credibilidade se mostraram inferiores a 1, indicando que o modelo proposto apresentou uma distribuição *a posteriori* de p mais concentrada sobre o parâmetro de interesse. Em outras palavras, apesar de as duas abordagens se manterem próximas em termos de distribuição *a posteriori*, o modelo FVO carregou uma quantidade mais significativa de informação inferencial sobre p que o modelo FVC, capacitando-o a apresentar estimativas e soluções mais precisas em um processo de tomada de decisões.

Numa segunda fase da tese, a função de verossimilhança ordenada foi novamente utilizada para atender a outro enfoque de interesse: a determinação do tamanho da amostra em experimentos binomiais. Embora esse tópico gere, a cada dia, mais e mais trabalhos e artigos, motivando os pesquisadores a apresentarem as mais diferentes soluções, não foram encontradas soluções para o problema específico desta tese. A inovação veio com o desejo de encontrar não só o tamanho da amostra (n) ótimo, mas também, e conjuntamente, o número de classificações repetidas (m) e o melhor critério classificador (a) de um item após terem sido realizadas m classificações repetidas do mesmo, com $1 \leq a \leq m$. Além disso, as componentes do vetor (n, m, a) ainda deveriam, simultaneamente, gerar uma variância *a posteriori* máxima inferior a uma variância σ^2 especificada e com uma amplitude máxima de intervalo de máxima densidade *a posteriori* (MDP) inferior a w , (também especificada) para uma cobertura mínima de $1-\alpha$.

A essa duas condições inferenciais chamam-se de funções objetivo e o problema é dito ser uma otimização combinatória multiobjetivo, pois o vetor (n, m, a) tem de atender a duas condições que envolvem diretamente as suas componentes. Então, utilizou-se uma abordagem multiobjetivo original.

Para a situação multiobjetivo, foi concebido um algoritmo evolutivo, denominado de algoritmo genético híbrido multiobjetivo (AGHM). Em sua construção, o AGHM envolveu um pouco dos conceitos de redes neurais artificiais e de busca local multiobjetivo, o que o transformou num programa computacional híbrido, mais robusto e mais eficiente. Do ponto de vista de simulação, foram analisados dois estudos desse algoritmo, uma versão mais rápida e outra mais precisa, cujos resultados se revelaram bastante animadores.

A versão mais rápida considerou uma Matriz-Trabalho de tamanho 600, analisando todos 24 pares (n, m) de pontos do espaço de respostas factíveis. Já o estudo 2, investigou um conjunto de

todos os pares ($n_{max} \leq 600$, $m_{max} \leq 7$) de pontos de soluções candidatas. Como resultado, o algoritmo forneceu soluções que combinou as grandezas (n , m , a), que são vitais para o planejamento e para o gerenciamento, fornecendo subsídios para análise de sensibilidade sobre os riscos (variabilidade) de projeto, bem como sobre os limites de especificação (comprimento do intervalo de credibilidade) de produtos.

Assim, um analista de produção ou um tomador de decisão, de posse de um conjunto de resultados como esse, tornaria as suas atitudes estratégicas de mercado em cima de uma lista de soluções boas para n , m e a conjuntamente.

É importante ressaltar que todos os códigos utilizados nessa tese foram construídos na plataforma do *Statistical Analysis System* – SAS, versão 9.1.3 e podem ser obtido solicitando-os diretamente ao autor por meio do endereço eletrônico joab@de.ufpb.br.

6.2. EXTENSÕES

Ainda assim, é possível melhorar, expandir e generalizar as metodologias discutidas nesse texto. Para o estudo do modelo preditivo FVO proposto, sugere-se reavaliá-lo, por exemplo, considerando os erros de classificação como variáveis latentes, semelhante a de um modelo de efeitos aleatórios, em que cada item avaliado teria um erro de classificação diferente, aleatório, mas próprio daquele item. Outra sugestão relacionada aos erros de julgamento é considerá-los como uma função contínua de m estritamente decrescente, de modo que os erros de classificação decaíssem à medida que aumentasse o número de classificações. É como se o processo classificatório sofresse uma espécie de aprendizagem, uma rede neural com método de aprendizagem por retropropagação, errando cada vez menos (aprendendo com os erros passados) à medida que o número de

classificações repetidas fosse aumentando. Assim, o aumento do parâmetro m faria com que, mais rapidamente, o processo de classificação não apresentasse quase erro algum. Também é possível uma generalização no sentido de que conjuntos de indivíduos ou itens diferentes tivessem quantidades de classificações repetidas diferentes.

Uma outra possível extensão é a generalização de dados binários para dados multinomiais, ou seja, uma nova função de verossimilhança poderia ser formulada considerando agora k categorias possíveis para a classificação de um determinado item e não mais somente duas.

Além disso, pode-se considerar não só dois níveis de informação para a distribuição *a priori* dos erros, mas uma família de distribuições *a priori*, como, por exemplo, uma família $\mathbf{B}=\{\text{Beta}(\alpha,\beta) ; \alpha,\beta \geq 1\}$ de distribuições *a priori* beta. E então essa informação, isto é, os parâmetros α e β da família \mathbf{B} poderiam ser incluídos no vetor de entrada (n, m, a, α, β) do algoritmo e, portanto, uma solução factível forneceria não só quais são os valores ótimos de n, m e a , mas também quais são os parâmetros α e β da melhor distribuição Beta que poderia ser usada, de modo que ajudasse a minimizar, simultaneamente, a variância *a posteriori* e o comprimento do intervalo de credibilidade.

E por fim, com relação à técnica de otimização multiobjetivo via algoritmos genéticos híbridos, uma sugestão, para construções futuras, é a confecção de outros sistemas de hibridação, como, por exemplo: (a) Lógica nebulosa Fuzzy com algoritmos genéticos; (b) Redes Neurais com Busca Tabu; (c) algoritmos genéticos com Colônia de Formigas. Esses e muitos outros processos de hibridização de algoritmos evolutivos podem ser alternativas eficientes para estudar tanto dados binomiais quanto outras estruturas numéricas, como dados multivariados, espaciais, de séries temporais, entre outros.

Como pode ser observado, o leque de novas investigações e pesquisas nessa área de análise bayesiana e em outras áreas de conhecimento é muito amplo e diversificado.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ACTON, F. S. **Numerical methods that works**. Mathematical Association of America, 3a. ed, Washington, DC, 1990.
- [2] AZZALINI, A. **A class of distributions which includes the Normal ones**. Scand J. Statistic, 12: pp. 171–178, 1985.
- [3] BARRY; CHRISTOPHER, B. **Bayesian evaluation of inspectors in sensory tests**. Journal of Quality Technology, 13, 2, pp. 120-124, 1981.
- [4] BERGER, J. O. **Statistical decision theory and bayesian analysis**, 2. ed, New York: Wiley, 1985.
- [5] BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: Wiley, 1992.
- [6] BROSS, I. **Missclassification in 2 x 2 tables**. Biometrics, vol. 10, pp. 478-486, 1954.
- [7] BURKE, J. R.; DAVIS, R. D.; KAMINSKY, F. C.; ROBERTS, A. E. P. **The effect of inspector errors on the true fraction non-conforming: An industrial experiment**. Quality Engineering, 7, 3, pp. 543-550, 1995.
- [8] COELLO, C. A. C. **An updated survey of ga-based multiobjective optimization techniques**. ACM Computing Surveys, vol. 32, 2, pp. 109-143, 2000.
- [9] DEB, K.; AGRAWAL, S.; PRATAB, A.; MEYARIVAN, T. **A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II**. KanGAL report, 2000.
- [10] DENDUKURI, N.; RAHME, E.; BÉLISLE, P.; JOSEPH, L. **Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test**. Biometrics, 60, pp. 388-397, 2004.
- [11] EHRGOTT, M. **Approximation algorithms for combinatorial multicriteria optimization problems**. International Transactions in Operational Research, vol. 7, pp. 5-31, 2000.
- [12] EVANS, M.; GUTTMAN, I.; HAITOVSKY, Y.; SWARTZ, T. **Bayesian analysis of binary data subject to misclassification**. In D. Berry, K. Chaloner, and J. Geweke (Eds.), Bayesian Analysis In Statistics and Econometrics: Essays In Honor Of Arnold Zellner, pp. 66–77. North Holland, New York, 1996.
- [13] EVANS, M.; GUTTMAN, I.; OLKIN, I. **Numerical aspects in estimating the parameters of a mixture of normal distributions**. Journal of Computing and Graphics Statistics, vol. 1, 4, pp. 351–365, 1993.
- [14] FLEISS, J. L. **Statistical methods for rates and proportions**. John Wiley and Sons, New York, 1973.
- [15] FUJISAWA, H.; IZUME, S. **Inference about misclassification probabilities from repeated binary responses**. Biometrics, 56, pp. 706-711, 2000.
- [16] GABA, A.; WINKLER, R. L. **Implication of errors in survey data: A bayesian model**. Management Science 38, 7, pp. 913–925, 1992.

- [17] GARCIA, J. J. *et al.* **Determination of IgE antibodies to the benzylpenicilloyl determinant: A comparison of the sensitivity and specificity of three radio allegro sorbent test methods.** *Journal of Clinical Laboratory Analysis* 11, pp. 251–257, 1997.
- [18] GELMAN, A. **Bayesian data analysis.** 2. ed. Chapman & Hall, London, 2004.
- [19] GOLDBERG, D. E. **Genetic algorithms in search, optimization and machine learning.** Reading, M.A.: Addison-Wesley, 1989.
- [20] GRAMOPADHYE, A. K.; PAJJURI, V.; KIMBLER, D. L.; FERRELL, W. G. **Compensating for inspection errors in attribute inspection.** *Quality Engineering*, 8, 2, pp. 311-322, 1995.
- [21] GREENBERG, B. S.; STOKES, S. L. **Repetitive testing in the presence of inspection errors.** *Technometrics*, 37, 1, pp. 102-111, 1995.
- [22] GUPTA, A. K.; NADARAJAH, S. **Handbook of Beta Distribution and Its Applications.** New York: Marcel Dekker, 2004.
- [23] GUSTAFSON, P. **Measurement error and misclassification in statistics and epidemiology: Impacts and bayesian adjustment.** Chapman & Hall, New York, 2003.
- [24] HAYKIN, S. **Redes neurais: princípios e prática.** tradução Paulo Martins Engel. 2. ed. Porto Alegre: Bookmann, 2001.
- [25] HOLLAND, J. H. **Adaptation in natural and artificial systems.** Cambridge, M.A.: MIT Press., 1975.
- [26] HYNDMAN, R. J. **Compute and graphing highest density regions.** *The American Statistician*, 50, pp. 120-126, 1996.
- [27] ISHIBUCHI, H.; MURATA, T. **A multi-objective genetic local search algorithm and its application to flowshop scheduling.** *IEEE Transactions on Systems, Man and Cybernetics-part C: Applications and Reviews*", vol. 28, pp. 392-403, 1998.
- [28] JEFFREYS, H. **Somes tests of significance, treated by the theory of the probability.** *Proceeding of the Cambridge Philosophy Society*, 31, pp. 203-222, 1935.
- [29] JEFFREYS, H. **Theory of probability.** Oxford: Claredon Press, 477p, 1961.
- [30] JOHNSON, N. L.; KOTZ, S. **Estimation from binomial data with classifiers of known and unknown imperfections.** *Naval Research Logistics*, 35, pp. 147–156, 1991.
- [31] JOHNSON, N. L.; KOTZ, S.; WU. X. **Inspection errors for attributes in quality control.** Chapman & Hall, London, 1991.
- [32] JOHNSON, S. L.; FUNKE, D. J. **An analysis of human reliability measures in visual inspection.** *Journal of Quality Technology*, 21, 2, pp. 71-74, 1980.
- [33] JONES, D. F.; MIRRAZAVI, S. K.; TAMIZ, M. **Multi-objective meta-heuristics: An overview of the current state-of-art.** *European Journal of Operational Research* vol. 137, pp. 1-19, 2002.
- [34] JOSEPH, L.; GYORKOS, T. W.; COUPAL, L. **Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard.** *American Journal of Epidemiology*, 141, 3, pp. 263-272, 1995.
- [35] JOSEPH, L.; WOLFSON, D. B.; DU BERGER, R. **Sample size calculations for binomial proportions via highest posterior density intervals.** *The Statistician*, 44, 2, pp. 143-154, 1995.

- [36] JOSEPH, L.; WOLFSON, D. B.; M'LAN, C. E. **Bayesian sample size determination for binomial proportions**. *Bayesian Analysis*, 3, 2, pp. 269-296, 2008.
- [37] KASS, R. E.; RAFTERY, A. E. **Bayes factors**. *Journal of the American Statistical Association*, vol. 90, 430, pp. 773–795, 1995.
- [38] KEMP, A. W.; KEMP, C. D. **A simple inspection scheme for two types of defect**. *Journal of the Operational Research Society*, 39, 3, pp. 311-315, 1988.
- [39] KLEINBAUM, D. G.; KUPPER, L. I.; MORGENSTERN, H. **Epidemiological research**. LifeTime Learning Publications, London. 1982.
- [40] KUTNER, M. H.; QU, Y.; TAN, M. **Random effects models in latent class analysis for avaluation accuracy of diagnostic tests**. *Biometrics*, 52, pp. 797–810, 1996.
- [41] LINDSEY, B. G. **Errors in inspection: integer parameter maximum likelihood in a finite population**. *Journal of the American Statistical Association*, 392, vol. 80, 1985.
- [42] LOSCHI, R. H., FRANCO, G. C. **Bayesian analysis for the meiosis I non-disjunction fraction in numerical chromosomal anomalies**. *Biometrical Journal*, 48, 2, pp. 220-232, 2006.
- [43] McCULLOCH, W. S.; PITTS, W. **A Logical calculus of the ideas immanent in nervous activity**. *Bulletin of Mathematical Biophysics*, 5, pp. 115-133, 1943.
- [44] MICHALEWICZ, Z. **Genetic algorithms+data+structures=evolution programs**. Springer-Verlag, 1994.
- [45] MORSE, J. N. **Reducing the size of the nondominated set: pruning by clustering**. *Computers and Operations Research*, 7, 2, pp. 55-66, 1980.
- [46] MOURA, F. M. **Avaliação de um sistema de inspeção dicotômico sem o conhecimento do estado real dos itens inspecionados**. 18 folhas. Dissertação (Mestrado em Estatística) – Instituto de Ciências Exatas – ICEx, Universidade Federal de Minas Gerais, Minas Gerais, 2007.
- [47] PIRES, M. C. **Análise bayesiana empírica de dados dicotômicos com erros e classificações repetidas**. 16 folhas. Dissertação (Mestrado em Estatística) – Instituto de Ciências Exatas – ICEx, Universidade Federal de Minas Gerais, Minas Gerais, 2006.
- [48] QUININO, R. C.; HO, L. L.; TRINDADE, A. L. G. **Bayesian judgment of a dichotomous inspection system when the true state of an evaluated item is unknown**. *Computers & Industrial Engineering*, 49, pp. 591-599, 2005.
- [49] RAHME, E.; JOSEPH, L.; GYORKOS, T. W. **Bayesian sample size determination for estimating binomial parameters from data subject to misclassification**. *Applied Statistic*, 49, 1, pp. 119-128, 2000.
- [50] ROSENBLATT, F. **The perceptron: A probabilistic model for information storage and retrieval in the brain**. *Psychological Review*, vol. 65, pp. 386-408, 1958.
- [51] RUBIN, D. B. **Using a SIR algorithm to simulate posterior distributions**. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 395–402. Cambridge, Oxford University Press, 1988.
- [52] RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning representations by back-propagation of errors**. *Nature*. 323, pp. 533-536, 1986b.

- [53] SCHAFFER, J. D. **Multiple objective optimization with vector evaluated genetic algorithms**. International Conference on Genetic Algorithms, Morgan Kaufmann, New York, pp. 93-100, 1985.
- [54] SRINIVAS, N.; DEB, K. **Multiobjective optimization using nondominated sorting in genetic algorithms**. Evolutionary Computation, vol. 2, 3, pp. 221-248, 1995.
- [55] STAMEY, J. D.; SEAMAN, J. W.; YOUNG, D. M. **Bayesian analysis of complementary poisson rate parameters with data subject to misclassification**. Journal of Statistical Planning and Inference, 134, pp. 36-48, 2004.
- [56] SYLLA, C. **Experimental investigation of human and machine-vision arrangements in inspection tasks**. Control Engineering Practice, 10, pp. 347-361, 2002.
- [57] TANG, M. L.; TANG, N. S.; CAREY, V. J. **Confidence interval for rate ratio in a 2x2 table with structural zero: An application in assessing false-negative rate ratio when combining two diagnostic tests**. Biometrics 60, pp. 550–555, 2004.
- [58] TENENBEIN, A. **A double sampling scheme for estimating from binomial data with missclassifications**. Journal of the American Statistical Association, 65, pp. 1350-1361, 1970.
- [59] TENENBEIN, A. **A double sampling scheme for estimating from binomial data with missclassifications: size sample determination**. Biometrics, 27, pp. 935-944, 1971.
- [60] TENENBEIN, A. **A double sampling scheme for estimating missclassified multinomial data with applications to sampling inspection**. Technometrics, 14, pp. 187-202, 1972.
- [61] VIANA, M. A. G.; RAMAKRISHNAN, V.; LEVY, P. S. **Bayesian analysis of prevalence from the results of small screening samples**. Communications in Statistics – V Theory and Methods, vol. 22, 2, pp. 575-585, 1993.
- [62] WINKLER, R. L. **Bayesian inference and decisions**. 2. ed. Probabilistic Publishing, London, 2003.
- [63] ZITZLER, E.; THIELE, L. **Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach**. IEEE Transactions on Evolutionary Computation, vol. 3, 4, pp. 257-271, 1999.