

RELATÓRIO TÉCNICO

Nº: DP/DEA – 39809/07

Nº DE PÁGINAS: 102

Nº DE ANEXOS: -

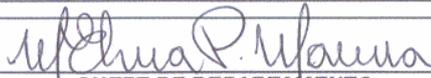
TÍTULO: Redução ótima de cenários em programação estocástica. Aplicação às vazões afluentes aos aproveitamentos hidroelétricos.**DEPARTAMENTO:** DEPARTAMENTO DE OTIMIZAÇÃO ENERGÉTICA E MEIO AMBIENTE - DEA

Nº DA ÁREA DE RESPONSABILIDADE: B200

Nº DA CONTA DE APROPRIAÇÃO: 1798

DESTINATÁRIOS:Às
Centrais Elétricas Brasileiras S/A - ELETROBRÁS
Av. Presidente Vargas, 409 – 12º andar
20071-003 – Rio de Janeiro / RJ**ATENÇÃO:** Paulo Roberto de Holanda Sales**EQUIPE DE ACOMPANHAMENTO:**Renata Leite Falcão
Arlete Rodarte Neves
Carlos Alberto de Barros Ferreira**RESUMO:**

Este relatório técnico apresenta o desenvolvimento teórico completo da técnica de redução ótima de cenários, os algoritmos de resolução, e aplicação desta em conjunto com o modelo de geração de séries sintéticas de vazões. Também é apresentada uma validação numérica com resultados promissórios para as usinas hidroelétricas que compõem o Sistema Interligado.

AUTOR(ES):Wellington Luis de Oliveira – CEPEL *wlo*
Claudia Sagastizabal – CEPEL *CS*
Maria Elvira Piñeiro Maceira – CEPEL *ME*
Débora Dias Jardim Penna – CEPEL *DDJP*
Jorge Machado Damázio – CEPEL *JMD***PALAVRAS-CHAVE:**Redução ótima de cenários
Geração sintética de cenários de vazões
Métrica de Fortet-Mourier**CLASSIFICAÇÃO:** CONTROLADO
GERENTE DO PROJETONOME: Débora Dias Jardim Penna
TEL: (21) 2598-6450 FAX: (21) 2598-6482
EMAIL: debora@cepel.br
CHEFE DE DEPARTAMENTONOME: Maria Elvira Piñeiro Maceira
TEL: (21) 2598-6454 FAX: (21) 2598-6482
EMAIL: elvira@cepel.br
DIRETOR DE PESQUISA E DESENVOLVIMENTO
ALBERT CORDEIRO GEBER DE MELO

12/12/2007

SUMÁRIO

RESUMO	1
1- INTRODUÇÃO AO PROBLEMA DE REDUÇÃO DE CENÁRIOS	2
1.1 – PRINCIPAIS CONTRIBUIÇÕES DO TRABALHO	4
1.2 – ESTRUTURA DO TEXTO	4
2 - UMA INTRODUÇÃO À PROGRAMAÇÃO ESTOCÁSTICA.....	6
2.1 – PROGRAMAÇÃO ESTOCÁSTICA EM DOIS E MULTI-ESTÁGIOS.....	7
2.1.1 – MODELOS COM RECURSOS EM DOIS-ESTÁGIOS	7
2.1.2 – MODELOS MULTI-ESTÁGIOS COM ALOCAÇÃO DE RECURSOS	11
2.2 – SÉRIE SINTÉTICA.....	14
3 - PROBLEMA DE TRANSPORTE DE MASSAS E RESULTADOS DE ESTABILIDADE	16
3.1 – TRANSPORTE DE MASSAS DE KANTOROVICH	17
3.2 – REPRESENTAÇÃO DUAL.....	19
3.2.1 – DUALIZAÇÃO DO PROBLEMA MK (CASO DISCRETO).....	19
3.2.2 – DUALIZAÇÃO DOS PROBLEMAS MK E KR - CASO CONTÍNUO	21
3.2.3 – RELAÇÃO ENTRE OS PROBLEMAS MK E KR	23
3.3 – MÉTRICAS PROBABILÍSTICAS E RESULTADOS DE ESTABILIDADE.....	26
3.3.1 – DISTÂNCIA VARIACIONAL	27
3.3.2 – DISTÂNCIA DE KOLMOGOROV-SMIRNOV (KS).....	27
3.3.3 – DISTÂNCIA DE FORTET-MOURIER	28
3.3.4 – DISTÂNCIA DE WASSERSTEIN.....	29
3.3.5 – RESULTADOS DE ESTABILIDADE	29
4 - O PROBLEMA DE REDUÇÃO DE CENÁRIOS	34
4.1 – O PROBLEMA	34
4.1.1 – REDUÇÃO ÓTIMA.....	35
4.1.2 – FORMULAÇÃO MATEMÁTICA	36
4.2 – ESCOLHA ÓTIMA DE CENÁRIOS	42
5 - ALGORITMOS DE SELEÇÃO E ALGORITMOS DE REDUÇÃO DE ÁRVORES DE CENÁRIOS.....	49

5.1 – ALGORITMOS PARA SELEÇÃO DE CENÁRIOS	49
5.2 – ALGORITMOS DE REDUÇÃO DE ÁRVORES DE CENÁRIOS	59
5.2.1 – REDUÇÃO E AGREGAÇÃO DE CENÁRIOS	60
5.2.2 – DIFICULDADES COMPUTACIONAIS	63
5.2.3 – REDUÇÃO LOCAL.....	64
<u>6 - RESULTADOS NUMÉRICOS.....</u>	68
6.1 – ANÁLISES ESTATÍSTICAS	69
6.1.1 – ÁRVORE DE CENÁRIOS UNIVARIADOS	69
6.1.2 – ÁRVORE DE CENÁRIOS MULTIVARIADOS.....	70
6.1.3 – CASO ESTUDADO	71
6.1.4 – REGRESSÃO LINEAR: MÉDIAS E DESVIOS PADRÕES	71
6.1.5 – REGRESSÃO LINEAR: CORRELAÇÃO TEMPORAL	72
6.1.6 – VALORES NUMÉRICOS DAS USINAS ESTUDADAS.....	73
6.1.7 – RESULTADOS - FIGURAS	76
6.2 – ANÁLISE DO DESEMPENHO DOS ALGORITMOS	86
6.2.1 – ANÁLISE DO ALGORITMO DE REDUÇÃO LOCAL.....	87
<u>7 - CONCLUSÕES.....</u>	90
<u>APÊNDICE A</u>	92
DUALIDADE EM PROGRAMAÇÃO LINEAR	92
<u>APÊNDICE B</u>	97
RELAÇÕES ENTRE OS MULTIPLICADORES DE LAGRANGE DO PROBLEMA DE MONGE-KANTOROVICH CASO DISCRETO	97
<u>REFERÊNCIAS BIBLIOGRÁFICAS</u>	100

RESUMO

O planejamento energético da operação de curto prazo é modelado como um problema de otimização estocástica multi-estágio. Variáveis incertas, como as vazões afluentes aos aproveitamentos hidroelétricos, são representadas por meio de uma árvore de cenários com probabilidades associadas. Porém, quando se utilizam árvores de mais de um estágio, a complexidade do problema de otimização resultante para o Sistema Interligado Nacional torna difícil a resolução em tempos computacionais razoáveis. Este trabalho estuda a técnica denominada de Redução Ótima de Cenários (ROC) para definir distâncias entre árvores de cenários, de maneira a obter árvores de portes menores que representem adequadamente o processo estocástico das afluições hidrológicas e sejam satisfatórias, tanto em termos de esforço computacional quanto em termos de qualidade da solução do problema de otimização. Mediante o uso da métrica probabilística de Fortet-Mourier e de resultados de dualidade em programação linear, a ROC permite calcular de forma explícita a distância entre duas medidas de probabilidade, correspondentes, respectivamente, à árvore de grande porte e a uma árvore de porte menor, com cardinalidade fixa. Por sua vez, esta distância explícita define a função objetivo para um outro problema de otimização, que seleciona os cenários mais representativos. Este problema de primeiro nível, de natureza combinatória, é de tipo *set-covering*. Para sua resolução é empregado um processo heurístico chamado de estratégia de redução máxima que fornece algoritmos rápidos de busca progressiva ou regressiva da cardinalidade da árvore reduzida.

O trabalho apresenta o desenvolvimento teórico completo da ROC, os algoritmos de resolução, e uma validação numérica com resultados promissórios para as usinas hidroelétricas que compõem o Sistema Interligado Nacional do Brasil.

1- Introdução ao Problema de Redução de Cenários

Diferentes campos de atuação envolvem estocasticidade: telecomunicações, finanças, transporte, meteorologia, entre outras, [23]. Deste modo, questões relacionadas às incertezas estão cada vez mais presentes nos setores industriais. Dentre as técnicas capazes de modelar as incertezas, pode-se destacar o importante papel exercido pela otimização estocástica. Um modelo de otimização estocástica tem por finalidade fornecer a melhor decisão a ser tomada na ocorrência de um cenário, ou realização particular do evento incerto analisado. Com o intuito de se representar melhor as incertezas, procura-se aumentar o número de cenários. Porém, devido aos altos tempos computacionais requeridos para se resolver o problema de otimização resultante, pode ser preferível trabalhar com um número menor de cenários, desde que sejam suficientemente representativos, [18]. Para atender estes dois interesses contraditórios pode se utilizar, por exemplo, a técnica de Redução Ótima de Cenários (ROC), assunto principal deste relatório.

Mais precisamente, este trabalho considera diferentes alternativas de redução de cenários, com aplicação para o problema do planejamento da operação energética do Sistema Interligado Nacional (SIN). Do ponto de vista da programação estocástica, trata-se de um problema de grande porte e complexo, devido à presença de um grande número de usinas hidroelétricas dispostas em cascata ao longo de uma mesma bacia hidrográfica. Por esta razão, no Brasil o problema do planejamento da operação é dividido em três etapas consecutivas – correspondentes ao médio e curto prazo, e a programação da operação – conforme apresentado em [19]. A cada uma das etapas correspondem modelos computacionais desenvolvidos pelo Centro de Pesquisas Elétricas (CEPEL), para a operação do SIN. Em particular, para o problema de planejamento da operação de curto prazo, o modelo DECOMP [19,27], determina a alocação ótima de recursos hídricos, térmicos e intercâmbios para cada estágio, minimizando o custo total esperado da operação ao longo do horizonte de planejamento, que pode ser de um ano. Neste contexto, as incertezas em relação às vazões afluentes aos aproveitamentos hidroelétricos são representadas por uma árvore de cenários. Esta árvore de cenários deve ser suficientemente representativa, do ponto de vista do processo estocástico das afluições, e ao mesmo tempo deve ser suficientemente compacta, de modo a garantir uma resolução do problema de otimização em tempos computacionais razoáveis. A técnica de ROC, estudada neste trabalho, visa atender estes dois importantes objetivos.

Em geral, dado um problema estocástico cuja incerteza é representada por uma árvore com um grande número de cenários, o problema de ROC consiste em achar um subconjunto reduzido de cenários mediante uma técnica que proporciona aderência estatística ao processo estocástico e decisões ótimas "semelhantes" às que se obteriam com o conjunto original. Especificamente, considere a função $f(w, x)$ definida no conjunto $\Theta \times X$, sendo Θ o espaço amostral onde as incertezas (cenários) variam, e X o conjunto

viável para a variável de decisão x^1 contidas em \mathbf{R}^n . Cada cenário possui uma probabilidade de ocorrência conhecida; a distribuição de probabilidades dos cenários é denotada por P . Para o problema

$$\min_{x \in X} \sum_{i=1}^N f(w_i, x) p(w_i),$$

procura-se reduzir o número de cenários para N_{red} , com N_{red} muito menor que N , e achar uma nova distribuição de probabilidades Q , tal que o subconjunto de cenários reduzidos mantenha alguns critérios de proximidade com relação ao conjunto original. Para a aplicação considerada, valores típicos de N e N_{red} são 10000 e 100, respectivamente. Uma característica importante da técnica ROC é que os N_{red} cenários escolhidos são, do ponto de vista do problema de otimização que se deseja resolver, os cenários mais representativos (este ponto será tratado em detalhes na Seção 4.1). Outro aspecto crucial do ponto de vista da implementação numérica é que a abordagem a priori mais natural, consistente em eliminar $N - N_{red}$ cenários por meio de análises exaustivas do conjunto de N cenários, pode não ser possível, tendo em vista a grande quantidade de combinações de conjuntos de cenários com cardinalidade $N - N_{red}$, i.e.,

$$\binom{N}{N - N_{red}}.$$

Deste modo, é preciso desenvolver técnicas heurísticas capazes de selecionar o melhor subconjunto de cenários, sem ter que percorrer todas as combinações possíveis. Estas técnicas devem também representar adequadamente as variações entre as soluções reduzidas (obtidas usando dados reduzidos) e as soluções originais (obtidas usando dados originais).

Para avaliar a proximidade entre os cenários de realização aleatória, são utilizadas métricas probabilísticas cujo desenvolvimento teórico parte do caso contínuo (correspondente a um número infinito de cenários). Tais métricas assumem um papel fundamental na ROC, pois uma boa escolha da métrica probabilística resulta em fórmulas explícitas para a seleção de um subconjunto de cenários e, posteriormente, para calcular a redistribuição de probabilidades Q [28].

¹ Para o problema do planejamento da operação energética de curto prazo, (Θ, x, f) correspondem, respectivamente, ao conjunto de cenários hidrológicos, alocação e custos de geração e déficit.

1.1 – Principais Contribuições do Trabalho

Este trabalho efetua a redução de cenários hidrológicos para o problema do planejamento da operação energética de curto prazo utilizando as técnicas gerais desenvolvidas nos artigos [6,8,10]. Todos os teoremas e proposições destes artigos foram sistematicamente estudados e reestruturados neste manuscrito. Para um melhor embasamento teórico, foi incluído um capítulo sobre os problemas de transporte (Capítulo 3), bem como a apresentação de algumas métricas probabilísticas. Também neste capítulo, a subseção "Dualidade do problema MK (caso discreto)" descreve minuciosamente o resultado de dualidade do problema de transporte de Monge-Kantorovich que proporciona, com algumas condições adicionais, a métrica probabilística sobre a qual se aplica todo o desenvolvimento teórico dos algoritmos de ROC.

Foram incluídas na parte teórica a Proposição 2 que sob condições específicas assegura a igualdade entre dois problemas de transportes, e a Proposição 3 que determina um limite superior para a diferença dos valores ótimos de problemas estocásticos utilizando distintas medidas de probabilidades. Foi desenvolvido o Algoritmo 7, denominado algoritmo de redução local, que constrói uma árvore de cenários através de sucessivas gerações e reduções de cenários, contornando assim a inviabilidade de armazenamento de um número de informações excessivo.

Vale ressaltar que os resultados apresentados em [8,10] são referentes à árvores univariadas. Em contraste, neste trabalho são apresentados testes de redução de cenários em árvores multivariadas, isto é, quando a árvore considerada no processo de ROC tem cenários matriciais ao invés de vetoriais, veja subseção 6.1.2. Esta situação é comumente encontrada no SIN, já que a árvore de cenários representa todas as usinas hidroelétricas do sistema. Em geral, as configurações de estudos do SIN consideram mais de 100 usinas. Por esta razão, diferentemente de [8,10], são utilizadas novas funções para se calcular as distâncias entre estes cenários matriciais, e foram incluídos novos testes² estatísticos para validar a técnica de ROC.

1.2 – Estrutura do Texto

Uma vez posicionada a ROC, exposto o campo de atuação, e definido o objetivo básico deste trabalho, cabe agora descrever em linhas gerais o que se pretende em cada capítulo.

O Capítulo 2, a seguir, contém uma descrição sucinta dos problemas estocásticos com recursos em dois ou mais estágios. Em particular, os problemas de multi-estágios evidenciam a dificuldade intrínseca de se lidar com informações de alta dimensionalidade, isto é, com um grande número de cenários. Este capítulo é baseado em notas de aulas ministradas pela Dra. Claudia Alejandra Sagastizábal, em agosto de 2005, no IMPA.

² Outros testes efetuados no subconjunto de cenários selecionados são, por exemplo, a verificação da preservação das correlações temporais e espaciais entre as usinas.

O Capítulo 3 apresenta o problema de transportes de massas de Monge-Kantorovich, bem como sua representação dual, que por sua vez, desempenha um papel fundamental para a ROC. São apresentados resultados de estabilidade para um conjunto de soluções do problema

$$\min_{x \in X} \int_{\Theta} f(w, x) dP.$$

O Capítulo 4 analisa em detalhes o conceito de ROC. É exposto o método pelo qual se faz a escolha dos índices dos cenários que serão descartados, bem como fórmulas explícitas de redistribuição de probabilidades para os dados preservados. Neste capítulo são enunciados e provados os principais teoremas, os quais proporcionam fórmulas compactas para os algoritmos de seleção de cenários.

O Capítulo 5 explica por extenso os algoritmos heurísticos de seleção de cenários. São apresentados exemplos numéricos para uma melhor compreensão dos procedimentos efetuados por cada algoritmo.

O Capítulo 6 contém resultados obtidos com a técnica ROC quando a metodologia é aplicada ao problema de reduzir dados de aflúências às usinas hidroelétricas do SIN.

O Capítulo 7 finaliza com as principais conclusões obtidas no decorrer deste trabalho.

O presente trabalho tem ainda dois apêndices onde são demonstrados alguns resultados matemáticos importantes que foram empregados durante o desenvolvimento teórico da ROC. O apêndice A contém resultados básicos de dualidade e condições de otimalidade em programação linear. A proposição sobre a qual se apresenta as relações entre os multiplicadores de Lagrange do problema de Monge-Kantorovich é demonstrada no Apêndice B.

2 - Uma Introdução à Programação Estocástica

Quando utilizados no campo industrial, muitos problemas de otimização podem ser formulados como programas estocásticos convexos, que geralmente são de grande porte. Em particular, no setor elétrico, um exemplo arquetípico é o problema de planejamento energético de curto prazo de sistemas hidrotérmicos, caracterizado pelas incertezas de afluições aos reservatórios de água (usinas hidroelétricas), pelas incertezas nos preços de combustíveis (usinas termoeletricas) e pelas incertezas com respeito ao mercado futuro de energia (demanda).

Em geral, a estratégia adotada para modelar um problema estocástico de operação é usar como função objetivo o valor esperado dos custos. Para isto, devem-se conhecer quais são as possíveis realizações futuras (cenários) de todas as variáveis aleatórias envolvidas no problema, ou gerar estes cenários de alguma maneira sistemática. A cada possibilidade de ocorrência do evento aleatório é atribuída uma probabilidade

Dado $X \subset \mathbb{R}^n$, um conjunto não vazio, convexo³ e fechado; dado um espaço amostral $\Theta \subset \mathbb{R}^T$ fechado, tal que w é um evento de Θ – chamado cenário – considere-se a função de custos f , semicontínua inferiormente⁴ e convexa⁵ em relação a $x \in X$, dada por $f : \Theta \times \mathbb{R}^n \rightarrow \mathbb{R}$. Quando a cardinalidade de Θ é infinita, o valor esperado da função f em relação a uma distribuição de probabilidades $P \subset \Theta$, é definida como

$$E_P f(w, x) := \int_{\Theta} f(w, x) dP.$$

Quando Θ é um conjunto finito de N cenários,

$$E_P f(w, x) := \sum_{i=1}^N f(w_i, x) p_i.$$

Com esta notação, o programa estocástico de interesse pode ser escrito da seguinte forma:

$$\min_{x \in X} E_P f(w, x). \tag{2.1}$$

³ Seja C um subconjunto de \mathbb{R}^n , diz-se que C é convexo se para quaisquer $x \in C$, $y \in C$ e $\alpha \in [0, 1]$, $\alpha x + (1 - \alpha)y \in C$.

⁴ $X \subset \mathbb{R}^n$, uma função $f : X \rightarrow \mathbb{R}$ é dita semicontínua inferiormente no ponto $a \in X$ quando, para cada $\varepsilon > 0$ dado, pode se obter $\delta > 0$, tal que $x \in X$, $\|x - a\| < \delta \Rightarrow f(x) < f(a) + \varepsilon$. Diz-se que f é semicontínua inferiormente quando ela o é em todos os pontos de X .

⁵ Seja $C \subset \mathbb{R}^n$ um conjunto convexo, diz-se que a função $g : C \rightarrow \mathbb{R}$ é convexa em C quando para quaisquer $x \in C$, $y \in C$ e $\alpha \in [0, 1]$, tem-se $g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y)$.

Devido à característica citada de existir um grande número de cenários, na maioria dos problemas reais têm-se um grande número de variáveis envolvidas. Com isto, em grande parte dos casos, a resolução direta do problema é computacionalmente inviável, sendo necessária a aplicação de técnicas especiais para reduzir a complexidade. Uma destas técnicas consiste em decompor o problema tratado em problemas menores e mais simples, que são chamados de problemas locais, seja através da decomposição por preços ou por recursos. No entanto, há situações em que tais decomposições não são suficientemente eficazes, já que não conseguem contornar a inviabilidade computacional do problema decomposto. Neste caso, uma alternativa é a redução dos dados incertos, isto é, reformular o problema desconsiderando alguns possíveis eventos e descartando cenários $w \in \Theta$. Esta alternativa será o principal assunto dos próximos capítulos.

A formulação (2.1) abrange os chamados programas estocásticos com recurso em dois-estágios ou multi-estágios apresentados a seguir.

2.1 – Programação Estocástica em Dois e Multi-Estágios

A resolução de problemas de otimização estocástica com recursos se baseia na idéia de trabalhar o problema em uma forma decomposta, fixando valores das variáveis de decisão seqüencialmente, segundo as incertezas vão sendo reveladas com o decorrer do tempo, por estágios.

2.1.1 – Modelos com Recursos em Dois-Estágios

Nesta formulação, os dados estão reorganizados em dois níveis diferentes. Os dados correspondentes às variáveis de primeiro estágio são determinísticos: $f_1: \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ é uma função de custos convexa, as restrições lineares envolvem o vetor $b_1 \in \mathbb{R}^{m_1}$ e as matrizes $A_1 \in \mathbb{R}^{m_1 \times n_1}$ e $D \in \mathbb{R}^{m_1 \times n_1}$. Já as variáveis de segundo estágio dependem de dados incertos, variando com cada cenário $w \in \Theta$. Assim sendo, a função de custos convexa de segundo estágio é denotada por $f_2: \Theta \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ e para cada $w \in \Theta$ têm-se os vetores $b(w) \in \mathbb{R}^{m_2}$, $b_2(w) \in \mathbb{R}^{m_2(w)}$, e as matrizes $A_2(w) \in \mathbb{R}^{m_2(w) \times n_2}$ e $W(w) \in \mathbb{R}^{m_1 \times n_2}$.

Um programa estocástico com recursos em dois-estágios pode ser representado por:

$$\left\{ \begin{array}{l} \min_{x^1, \{x^2(w), w \in \Theta\}} f_1(x^1) + E_p f(w, x^2(w)) \\ s.a : \quad A_1 x^1 = b_1 \\ \quad \quad A_2(w) x^2(w) = b_2(w), \\ \quad \quad D x^1 + W(w) x^2(w) = b(w). \end{array} \right. \quad (2.2)$$

A variável x^1 representa o vetor de decisões anteriores à realização do evento incerto correspondente a algum cenário w . Este primeiro período, em que não há incertezas, é denominado de "primeiro estágio" e todas as variáveis relacionadas a este período são denominadas variáveis de primeiro estágio, bem como as restrições. Já o segundo estágio é determinado pelo conjunto de períodos em que há incertezas em uma ou mais variáveis. As variáveis de segundo estágio são dadas pelos vetores $x^2(w)$, sendo $w \in \Theta$ um cenário, isto é, uma possível realização dos dados incertos para os quais se deseja otimizar a variável $x^2(w)$.

Uma representação equivalente do problema (2.1) envolve a função valor $v_2(x^1)$, definida da seguinte forma:

$$v_2(x^1) := \begin{cases} \min_{\{x^2(w), w \in \Theta\}} & E_p f(w, x^2(w)) \\ \text{s.a:} & A_2(w)x^2(w) = b_2(w), \\ & W(w)x^2(w) = b(w) - Dx^1. \end{cases} \quad (2.3)$$

Uma vez conhecido o cenário w , a decisão de recurso do segundo estágio é o ponto $x^2(w)$, solução de $v_2(x^1)$. Neste contexto, x^1 se comporta como um parâmetro do problema (2.3).

Assim sendo, o problema (2.2) pode ser dado da seguinte forma:

$$\begin{cases} \min_{x^1} & f_1(x^1) + v_2(x^1) \\ \text{s.a:} & A_1 x^1 = b_1. \end{cases} \quad (2.4)$$

A proposição a seguir analisa algumas propriedades da função valor v_2 .

Proposição 1 *Seja $h: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e coerciva⁶, u um parâmetro dado em \mathbb{R}^p , $W \in \mathbb{R}^{m \times n}$ uma matriz de posto completo, $b \in \mathbb{R}^m$, e $D \in \mathbb{R}^{m \times p}$. Defina-se $v(u)$ como sendo:*

$$v(u) = \begin{cases} \min_{y \in \mathbb{R}^n} & h(y) \\ \text{s.a:} & Wy = b - Du. \end{cases} \quad (2.5)$$

A função v é convexa, e se $\lambda^(u)$ denota um multiplicador de Lagrange associado*

⁶ Uma seqüência $\{x_k\} \subset \mathbb{R}^n$ é crítica em relação a um conjunto C , se $\{x_k\} \subset C$ e $\|x_k\| \rightarrow \infty$ ou $\{x_k\} \rightarrow x$, para x na fronteira do conjunto C . Diz-se que a função $g: C \rightarrow \mathbb{R}$ é coerciva no conjunto C , quando para toda seqüência crítica $\{x_k\}$ em relação C , tem-se $\limsup_{k \rightarrow \infty} g(x_k) = \infty$.

à restrição $Wy = b - Du$ no ponto ótimo $y^*(u)$, isto é, satisfazendo $v(u) = h(y^*(u))$, então $\lambda^*(u)'D$ é um subgradiente da função v no ponto u ($\lambda^*(u)'D \in \partial v(u)$).

Prova. Como W possui posto completo, segue que o sistema linear $Wy = b - Du$, na variável $y \in \mathbb{R}^n$, sempre possui solução para todo $u \in \mathbb{R}^p$. Sendo a função h coerciva, segue-se que o par $(y^*(u), \lambda^*(u))$ sempre existe para qualquer que seja $u \in \mathbb{R}^p$.

Escrevendo a função Lagrangiana de (2.5) quando se relaxa a restrição $Wy = b - Du$:

$$L(y, \lambda) = h(y) + \lambda'(Wy - b + Du),$$

a função dual $\theta(\lambda) = \inf_y L(y, \lambda)$ satisfaz a propriedade chamada "dualidade fraca" que estabelece que $\theta(\lambda) \leq h(y)$ para todo $\lambda \in \mathbb{R}^m$ e para todo $y \in \mathbb{R}^n$. Em particular, para $y = y^*(u)$, $v(u) = h(y^*(u)) \geq \theta(\lambda)$, ou seja,

$$\begin{aligned} v(u) &\geq \inf_y \{h(y) + \lambda'(Wy - b + Du)\} \\ &= \inf_y \{h(y) + \lambda'(Wy - b + Du)\} + \lambda'(D\bar{u} - D\bar{u}), \end{aligned}$$

para \bar{u} um ponto qualquer em \mathbb{R}^p .

Portanto, $v(u) \geq \inf_y \{h(y) + \lambda'(Wy - b + D\bar{u})\} + \lambda'(Du - D\bar{u})$, para todo $\lambda \in \mathbb{R}^m$. Em particular, para $\lambda^*(\bar{u})$, o multiplicador de Lagrange ótimo associado a restrição $Wy = b - D\bar{u}$, vale que

$$\begin{aligned} v(u) &\geq \inf_y \{h(y) + \lambda^*(\bar{u})'(Wy - b + D\bar{u})\} + \lambda^*(\bar{u})'(Du - D\bar{u}) \\ v(u) &\geq v(\bar{u}) + \lambda^*(\bar{u})'D(u - \bar{u}), \end{aligned}$$

portanto, $\lambda^*(\bar{u})'D$ é um subgradiente de $v(\bar{u})$.

Para provar que $v(u)$ é convexa, considere o conjunto

$$M(u) = \{y \in \mathbb{R}^n : Wy = b - Du\}.$$

Como h é coerciva, dados $u_1, u_2 \in \mathbb{R}^m$ existem $y_1, y_2 \in \mathbb{R}^n$ tais que

$$\begin{aligned} v(u_1) &= h(y_1), Wy_1 = b - Du_1, \\ v(u_2) &= h(y_2), Wy_2 = b - Du_2. \end{aligned}$$

Dado $t \in [0,1]$, tem-se que

$$\begin{aligned} W[ty_1 + (1-t)y_2] &= tWy_1 + (1-t)Wy_2 \\ &= t(b - Du_1) + (1-t)(b - Du_2) \\ &= b - D[tu_1 + (1-t)u_2], \end{aligned}$$

ou seja, $ty_1 + (1-t)y_2 \in M(tu_1 + (1-t)u_2)$. Então,

$$\begin{aligned} v(tu_1 + (1-t)u_2) &= \min_{y \in M(tu_1 + (1-t)u_2)} h(y) \\ &\leq h(ty_1 + (1-t)y_2) \\ &\leq th(y_1) + (1-t)h(y_2) \\ &= tv(u_1) + (1-t)v(u_2), \end{aligned}$$

ou seja, a função v é convexa. ■

Quando o conjunto Θ é discreto e finito, temos que

$$E_p f_2(w, x^2(w)) = \sum_{w \in \Theta} p(w) f_2(w, x^2(w)),$$

sendo $p(w)$ a probabilidade de ocorrência do cenário w , para cada $w \in \Theta$. Neste caso, o cálculo de $v_2(x^1)$ pode ser decomposto em $\#\Theta$ subproblemas de otimização, segundo apresentado a seguir:

Escrevendo $v_2(x^1, w)$ como sendo a função valor na configuração do cenário w , tem-se que

$$v_2(x^1, w) = \begin{cases} \min_{x^2(w)} f_2(w, x^2(w)) \\ s.a.: A_2(w)x^2(w) = b_2(w) \\ W(w)x^2(w) = b(w) - Dx^1. \end{cases}$$

para cada $w \in \Theta$. Com esta notação, pode se escrever a função valor $v_2(x^1)$ como sendo

$$v_2(x^1) = \sum_{w \in \Theta} p(w) v_2(x^1, w).$$

Pela Proposição 1 e pela nova representação de $v_2(x^1)$, é fácil verificar que $v_2(\cdot)$ é convexa e que

$$\sum_{w \in \Theta} p(w) \lambda^*(x^1)' D \in \partial v_2(x^1).$$

Para resolver o problema de otimização (2.4) que é convexo e não diferenciável usam-se métodos de otimização especiais como planos cortantes e feixes (ver [1]).

Um exemplo típico de problema em dois-estágios é a expansão do sistema elétrico, onde as variáveis de primeiro estágio são referentes ao investimento, e as variáveis de segundo estágio são dadas pelas incertezas de demanda e hidrologia. Porém, para alguns problemas, as incertezas evoluem com o passar do tempo, e uma modelagem em dois-estágios não é mais adequada. A seguir consideram-se os chamados modelos multi-estágios.

2.1.2 – Modelos Multi-Estágios com Alocação de Recursos

Nos modelos com recursos em dois-estágios, todas as incertezas são resolvidas após passado o primeiro estágio de decisão; no entanto, em muitos problemas de otimização, as variáveis aleatórias são reveladas sequencialmente, com o desenrolar dos estágios, e as decisões devem ser analisadas em muitos períodos, sendo necessário um outro tipo de enfoque no tratamento das incertezas.

Os modelos multi-estágios têm a vantagem de evitar uma escolha "cega" no início do período de análise. Um aspecto chave é a evolução do fenômeno aleatório ao longo do tempo. Na formulação multi-estágio, o problema pode ter variável de primeiro estágio aleatória; neste sentido não há então distinção entre primeiro estágio e subseqüentes estágios.

Um problema de decisão num estágio t pode variar drasticamente dependendo das incertezas realizadas nos períodos anteriores. Por exemplo, para o problema de planejamento da operação energética de curto prazo, extremamente dependente das aflúncias aos reservatório, é natural utilizar um programa estocástico com multi-estágios. Além do mais, toda decisão a ser tomada, referente ao estágio t , leva em consideração o nível de água nos reservatórios até o final do estágio $t - 1$, deste modo, as decisões são acopladas no tempo.

Uma maneira de representar os possíveis estados das variáveis aleatórias nos modelos de multi-estágios é colocando os cenários em formato árvore, ver Figura 2.1. Cada cenário w_i tem T componentes (cada componente pode ser vetorial). É assumido que todos os cenários coincidem no instante de tempo $t = 0$, isto é, $w_1^0 = w_2^0 = w_3^0 = \dots = w_N^0$. Isto significa que $t = 0$ pode ser considerado como a raiz da árvore que contém N cenários. Cada nó da árvore corresponde à realização da variável aleatória para um estágio t . Vale ressaltar que um mesmo valor pode corresponder a diferentes cenários. Na Figura 2.1, por exemplo, o nó 4 pertence tanto ao cenário w_1 como ao cenário w_2 (ou seja, $w_1^2 = w_2^2$). Quando usado o formato árvore, a denominada multiplicidade indica quantos "ramos" da árvore são abertas em cada nó. Por exemplo, a árvore à direita na Figura 2.1, possui multiplicidade constante e igual a dois (árvore binária). A árvore à esquerda possui multiplicidade variável entre 1 e 2.

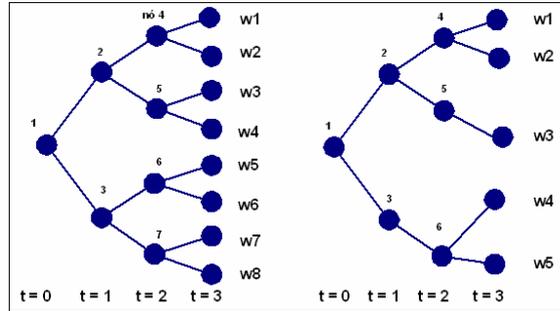


Figura 2.1: Cenários em formato árvore.

Devido ao acoplamento no tempo das variáveis decisão, tem-se a seguinte definição:

Definição 1 Implementabilidade

Um dos mais importantes aspectos no modelo com multi-estágios é a noção de implementabilidade ou "nonanticipativity" em inglês. Em geral, se dois cenários têm a mesma seqüência de nós entre o primeiro e o k-ésimo estágios, eles têm a mesma parcela de informações durante estes períodos. Conseqüentemente, decisões associadas a tais cenários devem ser idênticas até o k-ésimo período. Este requerimento é conhecido como condição de implementabilidade.

Dados T estágios de decisão com $T \geq 2$, para $t = 1, 2, \dots, T$, seja $X_t(w)$ o conjunto que contém todas as restrições do tipo $A_t(w)x^t(w) = \tilde{b}_t(w)$ generalizando as restrições do problema de segundo estágio (2.3), e também as restrições de implementabilidade descritas na Definição 1. No caso do modelo de dois-estágios, os conjuntos $X_1(w)$ e $X_2(w)$ representam, respectivamente, as restrições de primeiro estágio, ($X_1(w) \leftrightarrow A_1x_1 = b_1$) e de segundo estágio ($X_2(w) \leftrightarrow A_2(w)x^2(w) = b_2(w)$).

Existem também restrições acoplando dois estágios consecutivos, generalizando a restrição $W(w)x^2(w) = b(w) - Dx^1$ do problema (2.3). Assim, pode-se escrever o problema de multi-estágios da seguinte forma:

$$\left\{ \begin{array}{l} \min_x E_p \left[\sum_{t=1}^T f_t(w, x^t(w)) \right] \\ s.a: D_t(w)x^{t-1}(w) + W_t(w)x^t(w) = b_t(w) \quad w \in \Theta, t = 2, \dots, T \\ x^t(w) \in X_t(w) \quad w \in \Theta, t = 1, \dots, T. \end{array} \right. \quad (2.6)$$

O problema (2.6) é formulado por estágios de decisão "t". Para passar a uma formulação por nós, é conveniente abrir a árvore de cenários como mostra a Figura 2.2, onde cada linha vertical indica componentes idênticos dos cenários envolvidos, estas linhas

verticais representam a condição de implementabilidade dada na Definição 1.

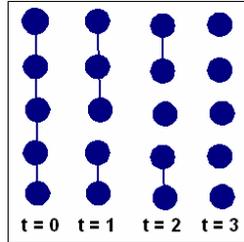


Figura 2.2: Restrições de implementabilidade

Na árvore da Figura 2.1, cada linha que liga o nó raiz a um dos nós do último estágio representa um evento aleatório, ou seja, um cenário. Logo esta árvore possui cinco cenários.

As linhas que ligam um nó a outro na Figura 2.2 representam as restrições de implementabilidade. Dos 20 nós existentes na Figura, há $4+3+2=9$ restrições. Resultando assim, $20-9=11$ nós como na árvore à direita na Figura 2.1. Por exemplo, se y^2 e y^3 representam decisões correspondentes a w^2 e w^3 , tem-se $y_3^2 = y_3^3$.

Devido a condição de implementabilidade, a cada estágio de tempo $t < T$, pode-se ter uma quantidade de decisões x_i estritamente menor que o número total de cenários N . Similarmente, pode-se chamar de A_t, D_t, W_t, \hat{b}_t e b_t os valores $A_t(w), D_t(w), W_t(w), \hat{b}_t(w)$ e $b_t(w)$ no estágio t , nos cenários w associados com o nó i , ver [24]. De forma análoga, introduzindo as funções de custo f_i , o problema (2.6) é equivalente ao problema, formulado por nós, dado pela expressão⁷

$$\begin{cases} \min_x \sum_{i \in \hat{N}} p_i f_i(x_i) \\ s.a : D_i x_{a(i)} + W_i x_i = b_i, i \in \hat{N} \setminus \{0\}, \\ A_i x_i = \hat{b}_i \end{cases} \quad (2.7)$$

com \hat{N} o conjunto de todos os nós da árvore. A função $a(i)$ representa o nó antecessor do i -ésimo nó (ver Figura 2.3).

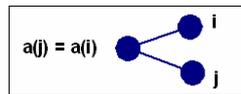


Figura 2.3: Nó antecessor do i -ésimo e j -ésimo nó.

⁷ As relações entre $A_t(w), D_t(w), W_t(w), \hat{b}_t(w), b_t(w)$, e $f_i(w, \cdot)$ do problema (2.6) e $A_i, D_i, W_i, \hat{b}_i, b_i$, e f_i do problema (2.7) vêm da passagem da árvore à direita na Figura 2.1 para a Figura 2.2.

A principal dificuldade no problema (2.7) é lidar com a dimensionalidade, seja de forma explícita, descrevendo os cenários no formato "árvore aberta" e usando as restrições de implementabilidade, ou de forma implícita, percorrendo os cenários de maneira encaixada, usando técnicas de programação dinâmica, por exemplo, de modo a se redefinir o problema original em subproblemas com dois-estágios.

O enfoque deste trabalho não é sugerir métodos de resolução de tais problemas, mas sim apresentar técnicas para reduzir a dimensionalidade dos cenários definindo o problema de otimização. O objetivo é tentar eliminar redundâncias e semelhanças de informações, reduzindo assim a dimensionalidade da árvore de cenários, de modo que a árvore reduzida obtida represente acuradamente o processo estocástico, e que seja viável em termos de tempo computacional. Assume-se aqui, "redundância" como sendo a ocorrência de eventos próximos, ou seja, eventos considerados "parecidos" segundo alguma métrica probabilística.

Para muitos problemas reais de otimização multi-estágios, os eventos aleatórios são gerados sistematicamente por modelos econométricos baseados em históricos de dados, como é o caso da análise de afluições aos reservatórios das usinas hidroelétricas.

A seguir é fornecido o conceito de séries sintéticas, uma forma de se trabalhar com cenários gerados a partir de modelos. Para mais informações, ver [7,14].

2.2 – Série Sintética

Freqüentemente, na impossibilidade de se conhecer as ocorrências futuras de fenômenos incertos, adota-se a hipótese de que ocorrências passadas, chamadas de "séries históricas", proporcionam uma descrição razoável do que pode acontecer no futuro. Por exemplo, quando se deseja planejar ou operar um aproveitamento hidroelétrico, a série histórica de vazões no local pode ser utilizada como dado de entrada para um modelo de otimização. Apesar de fornecer subsídios bastante úteis para algumas aplicações, os estudos com séries históricas apresentam limitações sérias, já que não é possível contemplar situações mais severas das registradas no histórico. Por este motivo, as séries históricas podem resultar insuficientes para compor uma amostra que estime acuradamente alguns índices de risco (como o não suprimento da demanda), associados com situações severas ou extremas. Já os modelos de geração de séries sintéticas de afluições a partir da série histórica resolvem a dificuldade mencionada, por serem capazes de capturar características básicas das ocorrências passadas, permitindo ao mesmo tempo a avaliação de riscos e incertezas extremos pertinentes a um sistema hidroelétrico, [18].

Denomina-se série sintética todo possível evento gerado por um modelo de uma variável aleatória, seja este físico, heurístico ou estocástico. Por exemplo, suponhamos que o vetor $w_i = (w_i^1, w_i^2, w_i^3, \dots, w_i^{T-1}, w_i^T)$ seja gerado a partir de um modelo. Esta é uma série sintética que assume valores w_i^t para todo instante de tempo $t=1,2,\dots,T$. Em geral, os modelos podem gerar tantas séries sintéticas quanto se queira: $w_i^t = \sum_{l=1}^p \Phi_i^{t-l} w_i^{t-l} + \xi_i^t$, sendo conhecidos os valores dos coeficientes Φ_i^t e das variáveis de estado w_i^{t-l} que foram

observados nos períodos anteriores; ξ_i^t é um ruído aleatório. Este modelo pertence a família dos modelos *ARIMA*, [2].

Considerando Ω como o conjunto de todas as possíveis séries sintéticas, é bastante intuitivo que Ω esteja contido no conjunto de todas as possíveis ocorrências do fenômeno aleatório considerado, isto é, que $\Omega \subset \Theta$. Logo, quanto maior for a quantidade de séries sintéticas geradas, maior é a chance de se estar próximo da incerteza representada por Θ . No entanto, a dimensionalidade de Ω pode deixar os programas de otimização computacionalmente inviáveis.

Uma possibilidade atrativa para escolher um subgrupo de séries sintéticas suficientemente representativas é utilizar como critério a medida probabilística chamada de "Fortet-Mourier", [22], que garante otimalidade em dois sentidos, descritos a seguir:

§ Estabilidade, no sentido que, para pequenas variações do conjunto original de séries sintéticas, resultem pequenas variações do subgrupo ótimo obtido.

§ Proximidade do grupo original, em termos de valores ótimos dos problemas de otimização correspondentes.

Também se espera que as propriedades acima auxiliem a boa aderência do grupo reduzido ao processo estocástico considerado. Os capítulos seguintes apresentam técnicas para se reduzir o número de séries sem que se percam informações estatisticamente significativas contidas no conjunto inicial (original) de séries sintéticas disponíveis.

3 - Problema de Transporte de Massas e Resultados de Estabilidade

Em termos mais precisos, se define a distância de Fortet-Mourier como o valor ótimo de um problema de transporte, [22]. A minimização desta distância sobre todos os possíveis subgrupos de séries sintéticas define o problema de ROC, assunto central deste trabalho.

Este capítulo se destina a uma introdução ao problema de transporte de massas de Monge-Kantorovich. O objetivo é apresentar as principais propriedades para então desenvolver a teoria da redução de cenários, dispostos no formato árvore.

As relações de dualidade do problema de transporte de massas são de grande relevância para as técnicas de seleção de cenários. Pode-se, por meio destas, escolher os cenários mais representativos do ponto de vista do problema de otimização que se deseja resolver.

Nas seções seguintes, são tratados dois tipos de problema de transporte: o problema de transporte de Monge-Kantorovich (MK), e o problema de transporte de Kantorovich-Rubinstein (KR).

Antes de descrever tais problemas, necessita-se das seguintes definições.

Definição 2 *Seja S um conjunto não-vazio. A função $d : S \rightarrow \mathbf{R}$ é dita ser uma métrica se para quaisquer dois pontos x, y em S , $d(x, y)$ é não-negativa e satisfaz as seguintes propriedades:*

- i) $d(x, y) = 0$ se, e somente se, $x = y$;
- ii) $d(x, y) = d(y, x)$ (simetria);
- iii) $d(x, y) \leq d(x, z) + d(y, z)$ (desigualdade triangular).

Definição 3 *Um conjunto S é dito ser um espaço métrico, se S é um espaço linear e possui uma métrica, $d : S \rightarrow \mathbf{R}$, associada⁸.*

Definição 4 *Um espaço S é dito ser separável, se S possui ao menos um subconjunto denso que seja enumerável.*

O espaço euclidiano \mathbf{R} , é um exemplo de espaço separável, pois \mathbf{Q} é um subconjunto denso e enumerável em \mathbf{R} . Munido da norma euclidiana, é também um espaço métrico.

⁸ Alternativamente, o espaço métrico associado à métrica d poderá ser representado por (S, d) .

3.1 – Transporte de Massas de Kantorovich

Dadas duas distribuições de probabilidades P e Q em um espaço métrico separável Θ , e dada uma função de custo μ definida no espaço amostral $\Theta_P \times \Theta_Q \subset \Theta \times \Theta$, o funcional de Kantorovich é definido por

$$MK(P, Q) := \begin{cases} \inf_{\eta} \int_{\Theta_P \times \Theta_Q} \mu(w, \tilde{w}) \eta(dP, dQ) \\ s.a.: \eta \in \Theta_P \times \Theta_Q \\ \eta(B \times \Theta_Q) = P(B) \\ \eta(\Theta_P \times B) = Q(B), \text{ para todo } B \subset \Theta_P \cap \Theta_Q. \end{cases} \quad (3.1)$$

Se as distribuições de probabilidades P e Q são discretas (com $\#P = N$ e $\#Q = M$), o problema (3.1) assume a seguinte forma,

$$MK(P, Q) = \begin{cases} \inf_{\eta} \sum_{i=1}^N \sum_{j=1}^M \mu(w_i, \tilde{w}_j) \eta_{ij} \\ s.a.: \eta_{ij} \geq 0, i = 1, \dots, N, j = 1, \dots, M \\ \sum_{i=1}^N \eta_{ij} = q_j, j = 1, \dots, M \\ \sum_{j=1}^M \eta_{ij} = p_i, i = 1, \dots, N, \end{cases} \quad (3.2)$$

sendo $q_j := Q(\tilde{w}_j)$ e $p_i := P(w_i)$ as probabilidades de ocorrência dos eventos $\tilde{w}_j \in \Theta_Q$ e $w_i \in \Theta_P$, respectivamente.

A Figura (3.1) proporciona uma intuição geométrica deste funcional.

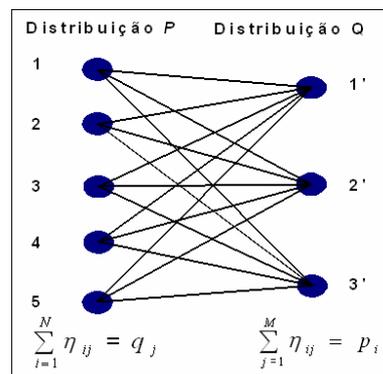


Figura 3.1: Representação do problema de Monge-Kantorovich.

Nas mesmas condições, define-se o problema de Kantorovich-Rubinstein como

sendo

$$KR(P, Q) := \begin{cases} \inf_{\eta} \int_{\Theta_P \times \Theta_Q} \mu(w, \tilde{w}) \eta(dP, dQ) \\ s.a: \eta \in \Theta_P \times \Theta_Q \\ \eta(B \times \Theta_Q) - \eta(\Theta_P \times B) = P(B) - Q(B). \end{cases} \quad (3.3)$$

Analogamente para o caso discreto,

$$KR(P, Q) = \begin{cases} \inf_{\eta} \sum_{i=1}^N \sum_{j=1}^M \mu(w_i, \tilde{w}_j) \eta_{ij} \\ s.a: \eta_{ij} \geq 0, i = 1, \dots, N, j = 1, \dots, M \\ \sum_{i=1}^N \eta_{ij} - \sum_{j=1}^M \eta_{ij} = q_j - p_i, i = 1, \dots, N, j = 1, \dots, M. \end{cases} \quad (3.4)$$

As distribuições de probabilidades P e Q dos problemas (3.1) a (3.4) podem ser vistas como distribuições de massas inicial e final. Por exemplo, para o problema (3.2), suponha que p_i seja a quantidade de uma certa mercadoria situada na localidade w_i , e que q_j seja a quantidade de mercadoria que deve ser entregue na localidade \tilde{w}_j . Levando-se em consideração que cada localidade de recebimento \tilde{w} deva receber mercadorias de todas as localidades de entrega w (como mostra a figura 3.1), o valor $\mu(w_i, \tilde{w}_j)$ representa o custo de transporte da mercadoria pelo trajeto η_{ij} . A variável η pode ser entendida como sendo planos de transportes. Se o ínfimo na variável η é atingido, suponha que seja em $\bar{\eta}$, diz-se que $\bar{\eta}$ é o plano ótimo de transporte. A existência de um minimizador $\bar{\eta}$ é assegurada se o conjunto $\Theta_P \times \Theta_Q$ for compacto, e a função μ for semicontínua inferiormente.

Assumindo que $\int_{\Theta_P \times \Theta_Q} \mu(w, \tilde{w}) \eta(dP, dQ) < \infty$, e devido ao fato de que o conjunto viável definindo KR contém aquele definindo MK, segue a desigualdade

$$KR(P, Q) \leq MK(P, Q). \quad (3.5)$$

Sob condições específicas, é possível obter a igualdade em (3.5). Tais condições são referidas à função μ , e serão doravante apresentadas.

3.2 – Representação Dual

Será apresentada nesta seção uma descrição da dualidade dos problemas (3.1) e (3.3). Começando pelo caso onde as distribuições de probabilidades são discretas, i.e., o espaço amostral Θ possui um número finito de eventos. Nestas condições, a função objetivo nos problemas (3.2) e (3.4) é afim. Logo, utiliza-se a teoria de programação linear para estender as relações duais aos problemas (3.1) e (3.3).

Para uma representação dual do problema (3.2), basta escrevê-lo de maneira equivalente ao problema (A.1), veja apêndice A.

3.2.1 – Dualização do Problema MK (caso discreto)

Suponha que para todo evento w_i em Θ , as probabilidades P e Q estejam bem definidas. Suponha também que, de forma mais geral, o número de átomos da distribuição P é diferente do número de átomos da distribuição Q ($N \neq M$).

Reescrevendo a função objetivo do problema (3.2), tem-se que

$$\sum_{i=1}^N \sum_{j=1}^M \mu(w_i, \tilde{w}_j) \eta_{ij} = \sum_{j=1}^M \mu(w_1, \tilde{w}_j) \eta_{1j} + \sum_{j=1}^M \mu(w_2, \tilde{w}_j) \eta_{2j} + \dots + \sum_{j=1}^M \mu(w_N, \tilde{w}_j) \eta_{Nj}. \quad (3.6)$$

Sejam os vetores \bar{C}_k e $X_k \in \mathbb{R}^M$, definidos por

$$\bar{C}_k := \begin{bmatrix} \mu(w_k, \tilde{w}_1) \\ \mu(w_k, \tilde{w}_2) \\ \vdots \\ \mu(w_k, \tilde{w}_M) \end{bmatrix} \text{ e } X_k := \begin{bmatrix} \eta_{k1} \\ \eta_{k2} \\ \vdots \\ \eta_{kM} \end{bmatrix}.$$

Definindo os vetores \bar{C} e $X \in \mathbb{R}^{N \times M}$ como sendo

$$\bar{C} := \begin{bmatrix} \bar{C}_1 \\ \bar{C}_2 \\ \vdots \\ \bar{C}_N \end{bmatrix} \text{ e } X := \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix},$$

é fácil ver que a expressão (3.6) é equivalente ao produto de \bar{C} por X :

$$\sum_{i=1}^N \sum_{j=1}^M \mu(w_i, \tilde{w}_j) \eta_{ij} = \bar{C}'_1 X_1 + \bar{C}'_2 X_2 + \dots + \bar{C}'_N X_N = \bar{C}' X.$$

$$\begin{cases} \max_{\Lambda} & b' \Lambda \\ \text{s.a:} & A' \Lambda \leq \bar{C}. \end{cases} \quad (3.8)$$

Assim sendo,

$$b' \Lambda = u' p + v' q = \sum_{j=1}^N u_j p_j + \sum_{j=1}^M v_j q_j,$$

e

$$A' \Lambda \leq \bar{C} = \begin{cases} u_1 + v_1 & \leq \mu(w_1, \tilde{w}_1) \\ & \vdots \\ u_1 + v_M & \leq \mu(w_1, \tilde{w}_M) \\ u_2 + v_1 & \leq \mu(w_2, \tilde{w}_1) \\ & \vdots \\ u_N + v_M & \leq \mu(w_N, \tilde{w}_M). \end{cases}$$

Logo, resolver o problema (3.2) equivale a resolver o problema (3.8), que por sua vez resulta em

$$\gamma_{\mu}(P, Q) = \max_{u, v} \left\{ \sum_{i=1}^N u_i p_i + \sum_{j=1}^M v_j q_j : u_i + v_j \leq \mu(w_i, \tilde{w}_j) \right\} \quad (3.9)$$

para $i=1, \dots, N$ e $j=1, \dots, M$. A extensão deste resultado ao caso contínuo se dá sem grandes complicações na subsecção a seguir.

3.2.2 – Dualização dos Problemas MK e KR - caso contínuo

A extensão das relações duais do problema MK ao caso contínuo, segue da definição de integrabilidade a Riemann. A integral de Riemann é definida como uma soma infinita de partições do domínio da função. Para fixar as idéias, na reta real, a integral definida de uma função é dada por

$$\lim_{n \rightarrow \infty} \sum_{i=0}^n m_i (t_{i+1} - t_i) \stackrel{b}{=} \int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=0}^n M_i (t_{i+1} - t_i),$$

onde o conjunto $\{t_0, t_1, \dots, t_n\}$ representa uma partição do intervalo $[a, b] \subset \mathbb{R}$. Os valores m_i e M_i são, respectivamente, o ínfimo e o supremo da função f no intervalo de reta $[t_i, t_{i+1}]$.

Desta maneira, podemos considerar o problema (3.1) como um programa linear (considerando uma partição suficientemente grande do domínio da função $\mu(\cdot; \cdot)$).

Procedendo como no caso discreto, a representação dual do problema (3.1) segue da forma

$$\gamma_{\mu}(P, Q) = \sup_{u, v} \left\{ \int_{\Theta_P} u(w) dP + \int_{\Theta_Q} v(\tilde{w}) dQ : u(w) + v(\tilde{w}) \leq \mu(w, \tilde{w}), w \in \Theta_P, \tilde{w} \in \Theta_Q \right\} \quad (3.10)$$

Para obter a representação dual do problema KR, é suficiente seguir os mesmos passos utilizados para a obtenção do dual do problema MK. Visto que o problema (3.4) é um programa linear (PL), e que (3.3) pode ser aproximado por um PL (devido à definição de integrabilidade à Riemann) a representação dual do problema KR segue da forma

$$\xi_{\mu}(P, Q) = \sup_u \left\{ \int_{\Theta} u(w) d(P - Q) : u(w) - u(\tilde{w}) \leq \mu(w, \tilde{w}), w, \tilde{w} \in \Theta \right\}. \quad (3.11)$$

Para mais detalhes, ver [22].

Se a função μ for simétrica, segue das restrições do problema (3.11) que

- i) $u(w) - u(\tilde{w}) \leq \mu(w, \tilde{w})$ para todo $w, \tilde{w} \in \Theta$;
- ii) $u(\tilde{w}) - u(w) \leq \mu(\tilde{w}, w)$;
- iii) $u(w) - u(\tilde{w}) \geq -\mu(\tilde{w}, w)$.

Usando a simetria da função μ , de i) e iii), tem-se que

$$-\mu(\tilde{w}, w) \leq u(w) - u(\tilde{w}) \leq \mu(w, \tilde{w}),$$

que é equivalente a

$$|u(w) - u(\tilde{w})| \leq \mu(w, \tilde{w}).$$

Logo, pode-se escrever (3.11) como

$$\xi_{\mu}(P, Q) = \sup_u \left\{ \int_{\Theta} u(w) d(P - Q) : |u(w) - u(\tilde{w})| \leq \mu(w, \tilde{w}), w, \tilde{w} \in \Theta \right\}.$$

Afirmção: O valor $\xi_{\mu}(P, Q)$ é não-negativo.

Suponha o contrário, i.e., $\int_{\Theta} u(w) d(P - Q) < 0$ para uma função \bar{u} solução do problema (3.11). Tomando $v := -\bar{u}$, tem-se que $\int_{\Theta} v(w) d(P - Q) > 0$, e as restrições $|v(w) - v(\tilde{w})| \leq \mu(w, \tilde{w})$, $w, \tilde{w} \in \Theta$ continuam sendo satisfeitas. Isto contradiz a hipótese de

que \bar{u} seja uma solução do problema (3.11).

Portanto não há alteração na solução do problema em considerar o valor absoluto na função objetivo do problema KR.

$$\Xi_{\mu}(P, Q) = \xi_{\mu}(P, Q) = \sup_u \left\{ \left| \int_{\Theta} u(w) d(P - Q) \right| : |u(w) - u(\tilde{w})| \leq \mu(w, \tilde{w}), w, \tilde{w} \in \Theta \right\}. \quad (3.12)$$

A seguir, algumas relações entre os funcionais de Monge-Kantorovich e Kantorovich-Rubinstein.

3.2.3 – Relação entre os problemas MK e KR

Serão apresentadas nesta subseção, algumas condições que asseguram a igualdade entre os funcionais (3.1) e (3.3). Primeiramente, suponha que a função μ satisfaça as condições de "função distância" seguintes

- i) $\mu(w, \tilde{w}) = 0$ se, e somente se, $w = \tilde{w}$;
- ii) $\mu(w, \tilde{w}) = \mu(\tilde{w}, w)$ para todo $w, \tilde{w} \in \Theta$ (simetria);
- iii) $\sup\{\mu(w, \tilde{w}) : w, \tilde{w} \in \mathbf{B}, \|w - \tilde{w}\| \leq \delta\}$ tende a zero quando δ tende a zero, para todo subconjunto \mathbf{B} em Θ ;
- iv) Exista uma função $\tau : \Theta \rightarrow \mathbf{R}_+$ limitada em conjuntos limitados, tal que: $\mu(w, \tilde{w}) \leq \tau(w) + \tau(\tilde{w})$ para todo $w, \tilde{w} \in \Theta$.

A seguir analisa-se quando estas propriedades são satisfeitas.

* Se $\mu(w, \tilde{w})$ é uma norma, as quatro propriedades valem.

* Se $\mu(w, \tilde{w})$ é uma função contínua, a propriedade *iii* é satisfeita.

* Se Θ é um conjunto compacto e $\mu(w, \tilde{w})$ é uma função contínua, (*iv*) vale, já que toda função contínua em um conjunto compacto é limitada. Assim, existe $M \in \mathbf{R}$ tal que, $\mu(w, \tilde{w}) < M$ para todo $w, \tilde{w} \in \Theta$, fazendo $\tau(w) = \max\{1, \|w\|\}M/2$, segue o resultado.

Um exemplo de função μ satisfazendo tais propriedades é dado por $\mu_h : \Theta \times \Theta \rightarrow \mathbf{R}_+$ como

$$\mu_h(w, \tilde{w}) = \|w - \tilde{w}\| \max\{1, h(\|w - w_0\|), h(\|\tilde{w} - w_0\|)\}$$

onde w_0 é um elemento dado em Θ , e $h: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ é uma função contínua e não decrescente.

Claramente, a função μ_h satisfaz as propriedades *i* a *iii*. Já a propriedade *iv* é satisfeita para

$$\tau_h(w) = 2 \|w - w_0\| \max\{1, h(\|w - w_0\|)\}.$$

Para verificar esta afirmação, considere três possibilidades na operação do máximo definindo μ_h (os casos restantes são análogos).

Caso 1 Se $\max(1, h(\|w - w_0\|), h(\|\tilde{w} - w_0\|)) = 1$, então

$$\begin{aligned} \mu_h(w, \tilde{w}) &= \|w - \tilde{w}\| \leq \|w - w_0\| + \|\tilde{w} - w_0\| \leq 2\|w - w_0\| + 2\|\tilde{w} - w_0\| \\ &= \tau_h(w) + \tau_h(\tilde{w}), \text{ como se queria mostrar.} \end{aligned}$$

Caso 2 Se $h(\|w - w_0\|) > 1 > h(\|\tilde{w} - w_0\|)$, então

$$\begin{aligned} \mu_h(w, \tilde{w}) &= \|w - \tilde{w}\| h(\|w - w_0\|) \leq \|w - w_0\| h(\|w - w_0\|) + \|\tilde{w} - w_0\| h(\|w - w_0\|). \\ \text{Como } h \text{ é não-decrescente, } \|w - w_0\| &\geq \|\tilde{w} - w_0\|, \text{ pela hipótese do caso, segue que} \\ \mu_h(w, \tilde{w}) &\leq 2\|w - w_0\| h(\|w - w_0\|). \text{ Por hipótese e pela definição de } \tau_h, \\ \tau_h(w) + \tau_h(\tilde{w}) &= 2\|w - w_0\| h(\|w - w_0\|) + 2\|\tilde{w} - w_0\|. \text{ Portanto,} \\ \mu_h(w, \tilde{w}) &\leq \tau_h(w) + \tau_h(\tilde{w}). \end{aligned}$$

Caso 3 Se $h(\|w - w_0\|) > h(\|\tilde{w} - w_0\|) > 1$, o resultado é análogo ao caso 2.

Logo, está comprovado que μ_h satisfaz também a condição *iv*.

A seguinte proposição estabelece condições para garantir a igualdade entre os problemas duais (3.2) e (3.4).

Proposição 2 Considere o caso discreto em que o espaço amostral Θ_P contém o espaço amostral Θ_Q , isto é,

$$\Theta_P = \{w_1, \dots, w_N\} \text{ e } \Theta_Q = \{\tilde{w}_1, \dots, \tilde{w}_M\} = \{w_{j_1}, \dots, w_{j_M}\},$$

sendo $\{w_{j_1}, \dots, w_{j_M}\} \subset \Theta_P$. Considere também a função μ satisfazendo as condições *i*) e *ii*) de função distância. Com estas considerações o par ótimo (\bar{u}, \bar{v}) , associado ao problema (3.9), possui a seguinte relação:

$$\bar{u}_i = -\bar{v}_{j_k}, \text{ se } i = j_k.$$

Prova. Ver apêndice B.

Definindo $q_i := 0$ para $w_i \in \Theta_P \setminus \Theta_Q$, e utilizando o resultado da Proposição 2, pode-se então reescrever o problema (3.9) da forma

$$\gamma_\mu(P, Q) = \max_u \left\{ \sum_{i=1}^N u_i (p_i - q_j) : u_i - u_j \leq \mu(w_i, \tilde{w}_j), i = 1, \dots, N, j = 1, \dots, M \right\}.$$

Logo, se a função $\mu(\cdot, \cdot)$ satisfaz as condições *i*-*iv*), segue a igualdade

$$\gamma_\mu(P, Q) = \xi_\mu(P, Q),$$

ou de forma equivalente,

$$MK(P, Q) = KR(P, Q). \quad (3.13)$$

Se a função μ for simétrica ($\mu(w_i, \tilde{w}_j) = \mu(\tilde{w}_j, w_i)$), usando o resultado (3.12), o problema dual de (3.2) é dado por

$$\gamma_\mu(P, Q) = \max_u \left\{ E_P u(w) - E_Q u(\tilde{w}) : |u(w_i) - u(\tilde{w}_j)| \leq \mu(w_i, \tilde{w}_j) \right\}, \quad (3.14)$$

para $i = 1, \dots, N$ e $j = 1, \dots, M$, sendo $E_P u(w) = \sum_{i=1}^N u(w_i) p_i$.

A extensão deste resultado ao caso contínuo foi mostrada pelo matemático Leonid V. Kantorovich em 1942. Este será o assunto a seguir.

Teorema 1 (*Teorema de Kantorovich*). *Considere o funcional*

$$\kappa(P, Q) := \sup_{f \in F_\mu} \left| \int_{\Theta_P} f(w) dP - \int_{\Theta_Q} f(w) dQ \right|, \quad (3.15)$$

para $F_\mu := \{g : \Theta \rightarrow \mathbb{R}; |g(w) - g(\tilde{w})| \leq \mu(w, \tilde{w})\}$, e para P e Q duas medidas de probabilidades em um espaço métrico separável Θ que contém os espaços amostrais Θ_P e Θ_Q . Considere também uma função μ como definida em $\Theta \times \Theta \rightarrow \mathbb{R}_+$, satisfazendo as condições (*i* - *iv*).

Então,

$$MK(P, Q) = \kappa(P, Q).$$

Prova. A prova deste teorema é feita em duas etapas. A primeira consiste em provar a desigualdade $MK(P, Q) \geq \kappa(P, Q)$, e a segunda etapa consiste em provar a

desigualdade contrária. Será feita a seguir a prova da primeira etapa; a segunda desigualdade pode ser verificada em [22].

Dada $f \in F_\mu$, e uma probabilidade bivariada η com distribuições marginais P e Q , se tem que

$$\begin{aligned}
 \left| \int_{\Theta_P} f(w) dP - \int_{\Theta_Q} f(w) dQ \right| &= \left| \int_{\Theta_P} \left[\int_{\Theta_Q} f(w) \eta(dP, dQ) \right] - \int_{\Theta_Q} \left[\int_{\Theta_P} f(\tilde{w}) \eta(dP, dQ) \right] \right| \\
 &= \left| \int_{\Theta_P} \left[\int_{\Theta_Q} f(w) \eta(dP, dQ) \right] - \int_{\Theta_P} \left[\int_{\Theta_Q} f(\tilde{w}) \eta(dP, dQ) \right] \right| \\
 &= \left| \int_{\Theta_P \times \Theta_Q} (f(w) - f(\tilde{w})) \eta(dP, dQ) \right| \\
 &\leq \int_{\Theta_P \times \Theta_Q} |(f(w) - f(\tilde{w}))| \eta(dP, dQ) \\
 &\leq \int_{\Theta_P \times \Theta_Q} \mu(w, \tilde{w}) \eta(dP, dQ).
 \end{aligned}$$

Como f e η são arbitrárias, tomando o \sup_f no lado direito e tomando o \inf_η no lado esquerdo da desigualdade acima, tem-se que $\kappa(P, Q) \leq MK(P, Q)$; como se queria mostrar. ■

Para o caso particular da ROC, o resultado do teorema de Kantorovich já havia sido provado em (3.13) e (3.14).

A seguir serão apresentadas algumas métricas probabilísticas. Em particular, será estudada a métrica de Fortet-Mourier que, por suas características intrínsecas, possui uma importante relação com o problema (3.14).

3.3 – Métricas Probabilísticas e Resultados de Estabilidade

Esta seção apresenta diferentes exemplos para medir a distância entre duas distribuições de probabilidades. Tais funções são chamadas de métricas probabilísticas. Serão apresentados também alguns resultados de estabilidade que permitem analisar a aplicabilidade de uma métrica probabilística dada para a ROC.

A escolha de uma métrica adequada é de grande importância para a ROC, pois determina quais serão os cenários selecionados. Também como consequência da métrica escolhida, são obtidos os melhores pesos para os cenários preservados (os cenários mantidos após aplicar a estratégia de ROC), mediante uma redistribuição de probabilidades.

A seguir as definições de distâncias probabilísticas. Para maiores detalhes sobre este

assunto, consultar a referência [11].

Métricas Probabilísticas

Sejam P e Q duas medidas de probabilidades e, $f \in F_\mu$ uma função semicontínua inferiormente definida em $\Theta \times \mathbb{R}^n \rightarrow \mathbb{R}$, onde F_μ é uma família qualquer de funções convexas com respeito à variável de decisão $x \in \mathbb{R}^n$.

Uma métrica probabilística define uma distância entre distribuições de probabilidades. Esta distância pode ser baseada nos valores esperados correspondentes a uma família de custos F_μ :

$$d(P, Q) = \sup_{f \in F_\mu} \left| \int_{\Theta} f(w, x) dP - \int_{\Theta} f(w, x) dQ \right|.$$

A seguir, serão apresentadas diferentes famílias F_μ , bem como suas correspondentes métricas probabilísticas.

3.3.1 – Distância Variacional

Esta distância utiliza o conceito de função indicadora. A função indicadora $\mathbf{1}_A$, também conhecida como "indicador" de um conjunto $A \subset \Theta$, é uma função que assume o valor 1 em A e 0 no seu complemento.

$$\mathbf{1}_A(w) := \begin{cases} 1 & \text{se } w \in A \\ 0 & \text{se } w \notin A. \end{cases}$$

A distância associada à classe de funções indicadoras é conhecida como *distância variacional*. Matematicamente, ela assume a seguinte forma:

$$d_V(P, Q) = \sup_{w \in \Theta} \int_{\Theta} \mathbf{1}_A(w) d(P - Q)(w),$$

onde $d(P - Q)$ é a função de densidade de probabilidades correspondente a diferença das distribuições P e Q .

3.3.2 – Distância de Kolmogorov-Smirnov (KS)

A distância KS é definida como a maior diferença entre as distribuições de

probabilidades acumuladas correspondente a P e Q , mais precisamente, sejam P e Q distribuições de probabilidades e $w \in \Theta$ um evento. A *distância de Kolmogorov-Smirnov* é dada pela expressão:

$$d_{KS}(P, Q) = \sup_{w \in \Theta} \left| \int_{-\infty}^w dP(t)dt - \int_{-\infty}^w dQ(t)dt \right|.$$

Tanto a distância KS quanto a distância variacional medem a proximidade entre P e Q independentemente de qualquer problema de otimização. A seguir consideram-se distâncias que levam em conta funções de custos de problemas de otimização para medir a proximidade entre as duas probabilidades.

3.3.3 – Distância de Fortet-Mourier

Seja $\mu: \Theta \times \Theta \rightarrow \mathbb{R}_+$ dada por

$$\mu(w, \tilde{w}) = \|w - \tilde{w}\| \max\{1, \|w - w_0\|^{r-1}, \|\tilde{w} - w_0\|^{r-1}\}, \quad (3.16)$$

para w_0 um elemento fixado *a priori* em Θ , e $r \geq 1$ um número real fixo. Para o caso de probabilidades associadas com árvores de cenários, uma possível escolha de w_0 é o vetor formado pelos valores esperados da distribuição P em cada estágio de tempo,

$$w_0 = [E_P w^1, E_P w^2, \dots, E_P w^T], \text{ com } E_P w^t = \sum_{i=1}^N p(w_i^t) w_i^t,$$

ver Figura 3.2.

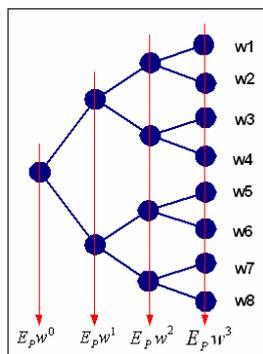


Figura 3.2: Uma possível escolha de w_0 .

Para esta função μ , a classe F_μ de funções Lipschitz-contínuas⁹ de ordem L , induz

⁹ Uma função f é Lipschitz-contínua com constante L se para quaisquer $x, y \in \mathbb{R}^n$ tem se que

a distância chamada de *Fortet-Mourier*:

$$d_{FM}(P, Q) = \sup_{f \in F_\mu} \left\{ \left| \int_{\Theta} f(w, x) dP - \int_{\Theta} f(w, x) dQ \right| : |f(w, \cdot) - f(\tilde{w}, \cdot)| \leq \mu(w, \tilde{w}) \right\}. \quad (3.17)$$

Note que a métrica d_{FM} nada mais é do que o funcional (3.15) definido no Teorema 1. Logo, resolver o problema MK equivale a medir a distância entre P e Q .

Desta forma, um motivo importante para se usar a métrica d_{FM} para selecionar cenários de uma árvore, é a preservação da proximidade entre os valores ótimos $E_P f(w, \bar{x})$ e $E_Q f(w, \tilde{x})$, onde $\tilde{x} \in B(\bar{x}, \varepsilon)$, para $\varepsilon > 0$ pequeno.

3.3.4 – Distância de Wasserstein

A distância de Wasserstein é um caso particular da distância de Fortet-Mourier. Mais precisamente, resulta de escolher $r = 1$ na função μ definida em (3.16):

$$d_w(P, Q) = \sup_{f \in F_\mu} \left\{ \left| \int_{\Theta} f(w, x) dP - \int_{\Theta} f(w, x) dQ \right| : |f(w, \cdot) - f(\tilde{w}, \cdot)| \leq \|w - \tilde{w}\| \right\}.$$

Pelos motivos mencionados nos itens *i)* e *ii)* acima, este trabalho faz uso da métrica de Fortet-Mourier. Passa-se agora a uma análise detalhada da função μ , fundamental para os resultados de estabilidade.

3.3.5 – Resultados de Estabilidade

Para uma medida de probabilidade dada, considere o problema (3.18) a seguir e os seguintes conjuntos

$$\min_{x \in X} E_P f(w, x) \quad (3.18)$$

$$V(P) := \inf \{ E_P f(w, x) : x \in X \} \text{ e} \quad (3.19)$$

$$S_\varepsilon(P) := \{ x \in X : E_P f(w, x) \leq V(P) + \varepsilon \}, \quad (3.20)$$

$$|f(x) - f(y)| \leq L \|x - y\|.$$

para cada $\varepsilon \geq 0$. Claramente, $V(P)$ é o valor ótimo do custo esperado para o problema (3.18) e $S(P) = S_0(P)$ é o conjunto solução.

Uma etapa fundamental da ROC consiste em determinar valores para uma distribuição de probabilidade Q , que mantenha a "proximidade" $V(Q)$ com $V(P)$. A noção de proximidade é avaliada através de resultados de estabilidade que proporcionam condições adequadas para a redistribuição dos cenários preservados, evitando a dispersão entre $S(Q)$ e $S(P)$.

Dada duas medidas de probabilidade P e Q , o Teorema 2 a seguir estima a proximidade entre $V(Q)$ ($S_\varepsilon(Q)$) e $V(P)$ ($S_\varepsilon(P)$), em relação a uma métrica probabilística dada. Para isto, considere a notação a seguir: $B(0, \rho)$ ¹⁰ é a bola fechada de raio ρ em \mathbb{R}^n , $\mathbf{P}(\Theta)$ uma família de subconjuntos de Θ , e \mathbb{P}_f o conjunto de medidas de probabilidades dado por:

$$\mathbb{P}_f := \left\{ Q \in \mathbf{P}(\Theta) : -\infty < \int_{\Theta} \inf_{x \in X \cap B(0, \rho)} f(w, x) dQ \leq \int_{\Theta} \sup_{x \in X \cap B(0, \rho)} f(w, x) dQ < \infty \text{ para todo } \rho > 0. \right\}$$

Considere também a distância entre $P, Q \in \mathbb{P}_f$

$$d_{f, \rho}(P, Q) = \sup_{x \in X \cap B(0, \rho)} \left| \int_{\Theta} f(w, x) dP - \int_{\Theta} f(w, x) dQ \right| \text{ para cada } \rho > 0.$$

Teorema 2 Dado $P \in \mathbb{P}_f$, suponha que $S(P)$ é não vazio e limitado. Então existem constantes $\rho > 0$ e $\bar{\varepsilon} > 0$ tais que

$$|V(P) - V(Q)| \leq d_{f, \rho}(P, Q) \text{ e } \emptyset \neq S(Q) \subset S(P) + M(d_{f, \rho}(P, Q))B(0, 1),$$

sempre que $Q \in \mathbb{P}_f$ com $d_{f, \rho}(P, Q) < \bar{\varepsilon}$. Além do mais, para qualquer $\varepsilon \in (0, \bar{\varepsilon})$ tem-se que:

$$d_\infty(S_\varepsilon(P), S_\varepsilon(Q)) \leq \frac{2\rho}{\varepsilon} d_{f, \rho+\varepsilon}(P, Q) \text{ quando } Q \in \mathbb{P}_f \text{ com } d_{f, \rho+\varepsilon}(P, Q) < \varepsilon.$$

Neste teorema, $M(\eta) := \eta + \psi^{-1}(2\eta)$, $\eta > 0$, onde

$$\psi(\tau) := \min_{x \in \mathbb{R}^n} \{E_p f(w, x) - V(P) : d(x, S(P)) \geq \tau\}, \text{ para } \tau > 0$$

¹⁰ Por definição, $B(0, \rho) := \{x \in \mathbb{R}^n : \|x\| \leq \rho\}$

é a função condicionada ao problema (3.18); $d_\infty(C, D) := \sup_{x \in \mathbb{R}^n} |d(x, C) - d(x, D)|$ é a distância de Pompeiu-Hausdorff entre os conjuntos fechados e não vazios $C, D \subset \mathbb{R}^n$; e $d(x, C) := \inf_{y \in C} \|x - y\|$ é a distância de $x \in \mathbb{R}^n$ a $C \subset \mathbb{R}^n$.

Para a demonstração, ver [6].

O Teorema 2 mostra que existe uma bola em \mathbb{R}^n , centrada em zero e com raio ρ , tal que a diferença em módulo dos valores esperados de duas distribuições de probabilidades com relação a função $f(\cdot, x)$ no ponto ótimo de (3.18) é menor que o supremo, também em módulo, dos valores esperados para todo x pertencente a esta bola:

$$|V(P) - V(Q)| \leq d_{f, \rho}(P, Q),$$

ou seja,

$$\left| \inf_{x \in X} E_P f(w, x) - \inf_{x \in X} E_Q f(w, x) \right| \leq \sup_{x \in X \cap B(0, \rho)} |E_P f(w, x) - E_Q f(w, x)|.$$

Além disso, o conjunto solução $S(Q)$ é não-vazio e está contido no conjunto solução $S(P) + B(0, \rho)$. Deste resultado segue que a diferença dos valores ótimos é finita, e que, se $S(Q)$ não está contido em $S(P)$, pelo menos está próximo deste conjunto.

O resultado observado vale quando \mathcal{P}_f é um subconjunto de $\mathbf{P}(\Theta)$ e $d_{f, \rho}(P, Q)$ é limitada superiormente por outra distância $d(P, Q)$ bem definida. A distância $d(P, Q)$ é chamada distância *canônica* ou *métrica probabilística ideal* associada ao problema (3.18), se ela tem a seguinte forma

$$d(P, Q) = \sup_{f \in \mathcal{F}_\mu |_\Theta} \left| \int f(w, x) dP - \int f(w, x) dQ \right|.$$

Um resultado de estabilidade envolvendo uma métrica ideal é apresentado na proposição seguinte.

Proposição 3 *Sejam $P, Q \in \mathcal{P}_f$ e considere que*

- a) O conjunto $S(P)$ é não vazio e $U \supseteq \mathbb{R}^n$ é uma vizinhança aberta e limitada de $S(P)$;
- b) $V_U(Q) := \inf \{E_Q f(w, x) : x \in X \cap \bar{U}\}$, $S_U(Q) := \{x \in X \cap \bar{U} : E_Q f(w, x) = V_U(Q)\}$;
- c) $F_\mu := \{f(\cdot, x) : x \in X \cap \bar{U}, |f(w, x) - f(\tilde{w}, x)| \leq \mu(w, \tilde{w}), \text{ para todo } w, \tilde{w} \in \Theta\}$;
- d) $\bar{x} \in S(P)$, $\tilde{x} \in S_U(Q)$, i.e, $E_P f(w, \bar{x}) = V(P)$ e $E_Q f(w, \tilde{x}) = V_U(Q)$.

Então, vale a seguinte desigualdade

$$|V(P) - V_U(Q)| \leq d(P, Q).$$

Prova. Sendo o conjunto U uma vizinhança do conjunto solução $S(P) \neq \emptyset$, e f um elemento da classe de funções F_μ , considere as possibilidades I e II.

$$\begin{array}{l}
 \text{I} \left\{ \begin{array}{l}
 0 \leq V(P) - V_U(Q) \\
 = E_P f(w, \bar{x}) - E_Q f(w, \bar{x}) \\
 \leq E_P f(w, \tilde{x}) - E_Q f(w, \tilde{x}) \\
 = \int_{\Theta} f(w, \tilde{x}) dP - \int_{\Theta} f(w, \tilde{x}) dQ.
 \end{array} \right.
 \quad
 \text{II} \left\{ \begin{array}{l}
 0 \leq V_U(Q) - V(P) \\
 = E_Q f(w, \tilde{x}) - E_P f(w, \bar{x}) \\
 \leq E_Q f(w, \bar{x}) - E_P f(w, \bar{x}) \\
 = \int_{\Theta} f(w, \bar{x}) dQ - \int_{\Theta} f(w, \bar{x}) dP.
 \end{array} \right.
 \end{array}$$

Logo,

$$\begin{aligned}
 |V(P) - V_U(Q)| &\leq \max \left\{ \int_{\Theta} f(w, \tilde{x}) dP - \int_{\Theta} f(w, \tilde{x}) dQ, \int_{\Theta} f(w, \bar{x}) dQ - \int_{\Theta} f(w, \bar{x}) dP \right\} \\
 &\leq \sup_{f \in F_\mu} \left| \int_{\Theta} f(w, x) dQ - \int_{\Theta} f(w, x) dP \right| \\
 &= d(P, Q),
 \end{aligned}$$

como se queria mostrar. ■

Considere agora o programa estocástico (3.18); assumindo que a função μ satisfaça as condições de função distância ($i - iv$), e que a distribuição de probabilidade P pertença a um espaço métrico separável Θ . Se o conjunto solução $S(P)$ de (3.18) é não vazio e limitado, pelo Teorema 2 e Proposição 3 podem ser verificadas propriedades de estabilidade nos valores $V(P)$ e nas soluções aproximadas $S_\varepsilon(P)$ com respeito a $d_{FM}(P, Q)$.

Como o teorema de Kantorovich garante a igualdade entre os funcionais $MK(P, Q)$ e $\kappa(P, Q)$, segue que $MK(P, Q) = d_{FM}(P, Q)$. Assim sendo, no caso discreto - o campo de interesse deste trabalho - o funcional de Kantorovich $MK(P, Q)$ (nestas condições, a métrica de Fortet-Mourier) será a principal ferramenta no processo de redução de cenários. Além disso, a métrica de Fortet-Mourier será de grande importância na redistribuição dos cenários preservados, pois através dela chega-se a uma fórmula de redistribuição que garante aderências estatísticas entre os dados originais e reduzidos, e deste modo, proporciona uma maior proximidade (medida pela métrica d_{FM}) entre os conjuntos soluções $S(P)$ e $S(Q)$.

Com os resultados apresentados é possível medir a proximidade entre os valores ótimos obtidos a partir dos dados originais e reduzidos. No entanto, ainda não foi mencionado nada a respeito de como se proceder para reduzir os cenários. Este assunto será

tratado no próximo capítulo, onde serão propostos e desenvolvidos métodos recursivos para a obtenção dos cenários a serem descartados, assim como fórmulas práticas para a redistribuição das probabilidades para os dados preservados.

4 - O Problema de Redução de Cenários

Este capítulo descreve os algoritmos necessários para realizar a redução ótima de cenários. Tais procedimentos são fundamentados nos resultados de estabilidade para métricas probabilísticas apresentadas no Capítulo 3, mais especificamente para a métrica de Fortet-Mourier, d_{FM} (Teorema 2 e relações (3.13) e (3.14)).

Como foi mencionado no Capítulo 2, muitos problemas de otimização podem ser modelados como programa estocástico em multi-estágios, no qual os eventos aleatórios são representados por uma árvore de cenários. A dimensionalidade destes cenários pode tornar o problema computacionalmente inviável, seja pela necessidade de um longo tempo de cálculo, ou até pela impossibilidade de processamento dos dados necessários para representar a árvore. Assim sendo, verifica-se a necessidade de reduzir o número de eventos para que seja possível resolver o programa estocástico.

4.1 – O Problema

Dado um conjunto de cenários organizados em formato árvore, seja P a distribuição de probabilidades discreta associada a este conjunto. Considere o problema de otimização

$$\min_{x \in X} E_P f(w, x) \quad (4.1)$$

onde $f: \Theta \times X \rightarrow \mathbb{R}$, $X \subset \mathbb{R}^n$ o conjunto das variáveis de decisão, $\Theta \subset \mathbb{R}^T$ o espaço amostral com N cenários w_i , $i \in I = \{1, 2, \dots, N\}$; cada cenários w_i com uma probabilidade $p_i = P(w_i)$ conhecida e $E_P f(w, x) = \sum_{i=1}^N f(w_i, x) p_i$ o valor esperado dos custos. Quando o número de cenários N é muito grande, por exemplo maior que 10000, a resolução direta de (4.1) é computacionalmente inviável. Para tanto, procura-se reduzir o número de cenários N para N_{red} , com N_{red} muito menor que N , de modo que a resolução do problema de otimização a seguir seja de resolução mais simples.

$$\min_{x \in X} E_{J^*, Q^*} f(w, x), \quad (4.2)$$

para $E_{J^*, Q^*} f(w, x) = \sum_{i \in I \setminus J^*} f(w_i, x) q_i$, onde $J^* = \{j_1, j_2, \dots, j_{N-N_{red}}\}$ é o conjunto ótimo de índices de cenários a serem descartados e $Q^* = \{q_1, q_2, \dots, q_{N_{red}}\}$ a redistribuição ótima dos cenários preservados. A escolha ótima do par (J^*, Q^*) será feita de modo a garantir a proximidade entre os problemas (4.1) e (4.2) (definida em (3.19), (3.20) e pelo Teorema 2).

4.1.1 – Redução Ótima

De modo geral, se procura obter um conjunto de índices $J^* = \{j_1, \dots, j_{N-N_{red}}\}$ e distribuição de probabilidades Q^* , tais que a distância entre os valores ótimos seja mínima

$$|V(P) - V(Q^*)| \leq \varepsilon \text{ e } S_\varepsilon(Q^*) \subset S_\varepsilon(P) + B(0, \rho),$$

para algum $\varepsilon \geq 0$ e $\rho \geq 0$ tão pequenos quanto possíveis.

Assim sendo, devem-se considerar os assuntos fundamentais a seguir:

- Como selecionar os índices de cenários a constituir o conjunto ótimo J^* que será descartado?
- Como obter a nova, e ótima, distribuição Q^* para os cenários preservados?

Intuitivamente, a resposta destas questões se baseia na minimização da distância entre os valores funcionais dos problemas (4.1) e (4.2):

$$\min_{J, Q} |E_P f(w, x) - E_{J, Q} f(w, x)| \quad (4.3)$$

Por meio desta representação é possível fazer uma analogia entre o problema (4.3) e a métrica de Fortet-Mourier definida em (4.4).

$$d_{FM}(P, Q) = \max_{f \in F_\mu} \left| \sum_{i=1}^N f(w_i, x) p_i - \sum_{j=1}^M f(x, \tilde{w}_j) q_j \right| \quad (4.4)$$

O uso da métrica de Fortet-Mourier é crucial para a ROC, pois a partir desta métrica pode-se obter, por dualidade em PL, uma representação explícita e compacta para determinar o par (J^*, Q^*) . Se a função μ satisfaz as condições de função distância ($i - iv$) do Capítulo 3, a métrica (4.4) coincide com o problema de transporte de Monge-Kantorovich¹¹, e a classe de funções F_μ é dada por

$$F_\mu = \{f(w_i, x) - f(\tilde{w}_j, x) \leq \mu(w_i, \tilde{w}_j), x \in X, i = 1, \dots, N, j = 1, \dots, M\}$$

Assim sendo, o par ótimo (J^*, Q^*) pode ser obtido por

¹¹ Note que esta métrica é a versão discreta do funcional definido no Teorema de Kantorovich 1.

$$\min_{J, Q} \{d_{FM}(P, Q) : \sum_{l \in I \setminus J} q_l = 1, q_l \geq 0, \#J = N - N_{red}\} \quad (4.5)$$

A minimização de d_{FM} , garante os resultados de estabilidade apresentados no Teorema 2.

A tarefa de selecionar os $N - N_{red}$ índices ótimos (que definem o conjunto J^*) pode ser formulada como um problema de programação inteira 0-1. A seguir, uma representação alternativa do problema de redução ótima de cenários (4.5).

4.1.2 – Formulação Matemática

Uma maneira alternativa de escrever o problema de ROC, consiste em separá-lo em dois níveis diferentes de otimização:

$$\begin{aligned} \min_J \{c(J) : \#J = N - N_{red}\}, \text{ onde} & \quad (\text{problema de primeiro nível}) \\ c(J) = \min_Q \{d_{FM}(P, Q) : \sum_{l \in I \setminus J} q_l = 1, q_l \geq 0\} & \quad (\text{problema de segundo nível}). \end{aligned}$$

O problema de segundo nível, $c(J)$, minimiza a métrica de Fortet-Mourier na variável Q , para um conjunto J fixo. A formulação da ROC em dois níveis tem o intuito de separar as variáveis e obter, para o problema de segundo nível (que define a função objetivo para o primeiro nível) uma formulação explícita, descrita a seguir. Tal formulação dispensa os algoritmos de programação inteira que, na maioria das aplicações exigem um esforço computacional muito elevado.

Representação Explícita do Problema de Segundo Nível.

O objetivo desta subseção é descrever uma representação compacta do problema $c(J)$, que não só servirá para encontrar a distribuição ótima Q^* , como também será utilizada para selecionar os índices de cenários $\{j_1, \dots, j_{N-N_{red}}\}$ a serem descartados.

Teorema 3 (Redistribuição Ótima). *Considere o problema $c(J)$; dado um conjunto $J \subset I = \{1, 2, \dots, N\}$, o valor mínimo do problema de segundo nível é dado por*

$$c(J) = \sum_{j \in J} p_j \min_{l \in I \setminus J} \mu(w_j, w_l). \quad (4.6)$$

Além disso, o mínimo é atingido em:

$$q_l = p_l + \sum_{j \in J_l} p_j \text{ para cada } l \in I \setminus J, \quad (4.7)$$

com $J_l := \{j \in J : l = l(j)\}$ e $l(j) \in \arg \min_{l \in I \setminus J} \mu(w_j, w_l)$ para cada $j \in J$.

Prova. Por conveniência notacional, seja $\mu_{ij} := \mu(w_i, w_j)$ para cada $i, j \in I$. Dado $J \subset I$, das representações primal (3.2) e dual (3.9) do problema $MK(P, Q)$ segue que

$$\begin{aligned} MK(P, Q) &= \min_{\eta} \left\{ \sum_{i,l} \mu_{il} \eta_{il} : \eta_{il} \geq 0, \sum_{i=1}^N \eta_{il} = q_l, l \in I \setminus J, \sum_{l \in I \setminus J} \eta_{il} = p_i, i \in I \right\} \\ &= \max_{u,v} \left\{ \sum_{i=1}^N u_i p_i + \sum_{l \in I \setminus J} v_l q_l : u_i + v_l \leq \mu_{il}, i \in I, l \in I \setminus J \right\} \end{aligned}$$

Lembrando que, pelo Teorema 1, $MK(P, Q) = d_{FM}(P, Q)$, ou equivalentemente, $\min_Q MK(P, Q) = c(J)$ para um conjunto J dado, faz-se a prova em duas partes.

-Parte 1: Será mostrado que $\sum_{j \in J} p_j \min_{l \in I \setminus J} \mu_{jl}$ é um limite inferior de $MK(P, Q)$ para u e v viáveis para a representação dual.

-Parte 2 : Será mostrado que $MK(P, Q) \leq \sum_{j \in J} p_j \min_{l \in J} \mu_{jl}$, sendo

$$Q^* = \{q_1^*, \dots, q_{N_{red}}^*\} \text{ os pesos ótimos de probabilidades.}$$

Parte 1: Sejam os vetores duais u e v com componentes $u_i := \min_{k \in I \setminus J} \mu_{ik}$ para cada $i \in I$, e $v_l = 0$ para cada $l \in I \setminus J$. Pela definição, $u_i = 0$ para cada $i \notin J$ e, portanto, para $l \in I \setminus J$ tem-se que

$$\begin{aligned} u_i + v_l &= u_i \leq \mu_{il} \quad \text{para todo } i \in J \\ u_i + v_l &= 0 \leq \mu_{il} \quad \text{para todo } i \in I \setminus J. \end{aligned}$$

Em conseqüência, u e v são viáveis para a formulação dual, e assim

$$\begin{aligned} \sum_{j \in J} p_j \min_{l \in I \setminus J} \mu_{jl} &= \sum_{i=1}^N u_i p_i + \sum_{l \in I \setminus J} v_l q_l \\ &\leq \max_{u,v} \left\{ \sum_{i=1}^N u_i p_i + \sum_{l \in I \setminus J} v_l q_l : u_i + v_l \leq \mu_{il}, i \in I, l \in I \setminus J \right\} \\ &= MK(P, Q). \end{aligned}$$

Logo, $\sum_{j \in J} p_j \min_{l \in I \setminus J} \mu_{jl} \leq MK(P, Q)$ para qualquer u e v viáveis, como se queria provar.

Parte 2: Seja o vetor primal $\bar{\eta}$ com componentes

$$\bar{\eta}_{il} := \begin{cases} p_i, & i \in J, l \in I \setminus J \\ p_i, & i = l \in I \setminus J \\ 0 & \text{caso contrário.} \end{cases}$$

Para cada $l \in I \setminus J$ definindo $q_l^* = \sum_{i=1}^N \bar{\eta}_{il}$ tem se que:

$$q_l^* = p_l + \sum_{i \neq l} \bar{\eta}_{il} = p_l + \sum_{i \in J_l} p_i.$$

Verifica-se também que

$$\sum_{l \in I \setminus J} \bar{\eta}_{il} = p_i,$$

tanto para $i = l$ e $l \in I \setminus J$, como para $i \neq l$ e $i \in J_l$. Como as componentes de $\bar{\eta}$ são todas não negativas, segue que $\bar{\eta}$ é viável para a representação primal de $MK(P, Q)$ quando $Q = Q^*$ e, portanto,

$$MK(P, Q) \leq \sum_{i,l} \mu_{il} \bar{\eta}_{il}.$$

Mas, para cada $i \in J$, pela definição de J_l , existe um $\bar{l} \in I \setminus J$ tal que $\mu_{i\bar{l}} := \min_{l \in I \setminus J} \mu_{il}$.

Assim sendo,

$$\sum_{i,l} \mu_{il} \bar{\eta}_{il} = \sum_{i \in J} p_i \mu_{i\bar{l}} = \sum_{i \in J} p_i \min_{l \in I \setminus J} \mu_{il},$$

como se queria provar. ■

Com este resultado crucial, a ROC passa a ter a seguinte estrutura:

$$\min_J \left\{ \sum_{i \in J} p_j \min_{l \in I \setminus J} \mu(w_j, w_l) : \#J = N - N_{red} \right\} \quad (4.8)$$

A correspondente redistribuição de probabilidades é dada por (4.7). Com esta formulação, a ROC consiste em conferir todos subconjuntos J de I com cardinalidade $N - N_{red}$; um problema de natureza combinatória que, dependendo do número de cenários, pode ser inviável computacionalmente. Mais precisamente, o número de candidatos J entre os quais se deseja escolher o conjunto J^* , é dado por:

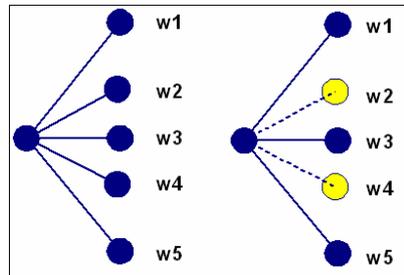
$$\binom{N}{N_{red}} = \frac{N!}{(N - N_{red})! N_{red}!}.$$

Se, por exemplo, a cardinalidade dos dados iniciais é $N = 100$, e deseja-se reduzir o número de cenários para $N_{red} = 40$, têm-se $1374.6 \cdot 10^{25}$ candidatos a J^* . Desta forma, calcular $c(J)$ para todo candidato J não é razoável. No próximo capítulo serão apresentados algoritmos, baseados nos valores de $c(J)$, que selecionam os melhores cenários a serem descartados. Estes algoritmos operam de forma "inteligente", proporcionando um esforço computacional bastante reduzido. O funcionamento de tais algoritmos será explicado através de exemplos.

Uma vez selecionados os índices do conjunto J^* , o Teorema 3 também apresenta uma fórmula explícita para se obter Q^* , a redistribuição ótima de probabilidades dos cenários preservados. Sua interpretação se resume ao fato de que a nova probabilidade para o cenário mantido deve ser igual a probabilidade que possuía, mais a soma de todas as probabilidades dos cenários próximos que foram descartados.

A fim de esclarecer o procedimento da regra ótima de redistribuição, o simples exemplo numérico a seguir explica a definição do conjunto $J_l = \{j \in J : l = l(j)\}$. Ao mesmo tempo, o exemplo permite a visualização da última passagem do Teorema 3, isto é, a igualdade $\sum_{i,j} c_{ij} \bar{\eta}_{il} = \sum_{i \in J} p_i \min_{l \in J} c_{il}$.

Exemplo 1 Considere os cenários representados na Figura 4.1a.



Figuras: a) Dados originais b) Dados reduzidos.

Suponha que se escolha $J = \{2,4\}$, ou seja, que os cenários 2 e 4 sejam descartados, conforme mostra a Figura 4.1b. Suponha ainda que para uma função μ dada, verifica-se que dentre os cenários preservados, o cenário 3 está mais próximo do cenário descartado 2: $\mu(w_2, w_3) \leq \mu(w_2, w_l)$ para todo $l \in I \setminus J$. Do mesmo modo, suponha que o cenário 3 está mais próximo do cenário descartado 4: $\mu(w_4, w_3) \leq \mu(w_4, w_l)$ para todo $l \in I \setminus J$.

Assim sendo,

$$3 = l(2) \in \arg \min_{l \in I \setminus J} \mu_{2l}, (\text{obs : } \mu_{ij} = \mu(w_i, w_j))$$

$$3 = l(4) \in \arg \min_{l \in I \setminus J} \mu_{4l}.$$

Segundo a definição de J_l , $J_3 = \{2,4\}$ e $J_1 = J_5 = \emptyset$, pois os cenários 1 e 5 não são os mais próximos dos cenários descartados.

Seja p_i a probabilidade de ocorrência do cenário w_i , para $i = 1,2,3,4,5$. Pela

definição de $\bar{\eta}_{il}$, segue o seguinte resultado para $l \in I \setminus J$:

$$\bar{\eta}_{11} = p_1 \quad i = j = 1,$$

$$\bar{\eta}_{23} = p_2 \quad i = 2 \in J_3,$$

$$\bar{\eta}_{33} = p_3 \quad i = j = 3,$$

$$\bar{\eta}_{43} = p_4 \quad i = 4 \in J_3,$$

$$\bar{\eta}_{55} = p_5 \quad i = j = 5; e$$

$\bar{\eta}_{il} = 0$, para todos outros casos.

Logo,

$$\begin{aligned} \sum_{i,l} \mu_{il} \bar{\eta}_{il} &= \mu_{11}p_1 + \mu_{23}p_2 + \mu_{33}p_3 + \mu_{43}p_4 + \mu_{55}p_5 \\ &= 0p_1 + \mu_{23}p_2 + 0p_3 + \mu_{43}p_4 + 0p_5 = \mu_{23}p_2 + \mu_{43}p_4 \\ &= p_2 \min_{l \in I \setminus J} \mu_{2l} + p_4 \min_{l \in I \setminus J} \mu_{4l} \\ &= \sum_{j \in J} p_j \min_{l \in I \setminus J} \mu_{jl}. \end{aligned}$$

As novas probabilidades dos cenários preservados são dadas por $q_l = p_l + \sum_{j \in J_1} p_j$.

Assim, a redistribuição é dada da seguinte forma:

$$Q^* = \{q_1 = p_1; q_2 = 0; q_3 = p_3 + p_2 + p_4; q_4 = 0; q_5 = p_5\}.$$

Podem-se obter outras formas de redistribuir probabilisticamente os cenários preservados. Uma alternativa é somar à p_l com $l \in I \setminus J$ apenas uma parcela da soma total das probabilidades p_j com $j \in J$, por exemplo, para cada $l \in I \setminus J$

$$q_l = p_l + \lambda_l p_J, \quad (4.9)$$

com $p_J = \sum_{j \in J} p_j$, $\lambda_l \geq 0, l \in I \setminus J$, e $\sum_{l \in I \setminus J} \lambda_l = 1$.

Quando se usa esta redistribuição, o teorema a seguir fornece um limite superior para $MK(P, Q)$.

Teorema 4 Quando o conjunto $J \subset \{1, 2, \dots, N\}$ é dado, e a probabilidade $Q = \{q_1, \dots, q_{N_{red}}\}$ é obtida pela regra de redistribuição (4.9), tem-se a seguinte desigualdade:

$$MK(P, Q) \leq \sum_{j \in J} p_j \sum_{l \notin J} \lambda_l \mu(w_j, w_l). \quad (4.10)$$

Além disso, vale a igualdade se $\#J = 1$ e se μ satisfaz a desigualdade triangular.

Prova. Usando a representação primal (3.2) de $MK(P, Q)$ e definindo a variável η como:

$$\eta_{jl} = \begin{cases} p_j \lambda_l & j \in J, l \in I \setminus J \\ p_j & j = l \in I \setminus J \\ 0 & \text{caso contrário,} \end{cases}$$

vale para qualquer $j \in I \setminus J$ que $\sum_{l \in I \setminus J} \eta_{jl} = \eta_{jj} = p_j$. Quando $j \in J$, novamente pela definição de η ,

$$\sum_{l \in I \setminus J} \eta_{jl} = p_j \lambda_{k_1} + p_j \lambda_{k_2} + \dots + p_j \lambda_{k_{N-N_{red}}} = p_j \sum_{i=1}^{N-N_{red}} \lambda_{k_i} = p_j,$$

e, portanto, $\sum_{l \in I \setminus J} \eta_{jl} = p_j$ para $j = 1, \dots, N$.

Assim sendo, η_{jl} satisfaz uma das restrições do problema (3.2). Quanto à restrição

$\sum_{i=1}^N \eta_{il} = q_l$, para $l \in I \setminus J$, tem-se que:

$$\begin{aligned} \sum_{i=1}^N \eta_{il} &= \underbrace{0+0+\dots+0}_{\text{para } i \notin J} + \underbrace{p_l}_{i=l} + \underbrace{p_{k_1} \lambda_l + p_{k_2} \lambda_l + \dots + p_{k_{N-n}} \lambda_l}_{i \in J} \\ &= p_l + \lambda_l \sum_{k \in J} p_k = p_l + \lambda_l p_J = q_l. \end{aligned}$$

Deste modo, todas as restrições estão satisfeitas já que, por definição, η_{jl} é não negativo. Logo, η_{jl} é viável para a representação primal de $MK(P, Q)$.

Pela definição de η , tem-se que

$$\begin{aligned} \sum_{j,l} \mu(w_j, w_l) \eta_{jl} &= \sum_l \mu(w_1, w_l) \eta_{1l} + \sum_l \mu(w_2, w_l) \eta_{2l} + \dots \\ &\quad + \sum_l \mu(w_N, w_l) \eta_{Nl}, \end{aligned}$$

sendo

$$\begin{aligned} \sum_l \mu(w_j, w_l) \eta_{jl} &= \begin{cases} 0, & \text{se } l, j \in I \setminus J \\ p_j \lambda_{k_1} \mu(w_j, w_{k_1}) + p_j \lambda_{k_2} \mu(w_j, w_{k_2}) + \dots \\ \quad + p_j \lambda_{k_{N_{red}}} \mu(w_j, w_{k_{N_{red}}}), & \text{se } l \in I \setminus J \text{ e } j \in J \end{cases} \\ &= p_j \sum_{l \in I \setminus J} \lambda_l \mu(w_j, w_l). \end{aligned}$$

Logo, $\sum_{j,l} \mu(w_j, w_l) \eta_{jl} = \sum_{j \in J} p_j \sum_{l \in I \setminus J} \lambda_l \mu(w_j, w_l)$ e (4.10) vale, pela definição de $MK(P, Q)$.

Finalmente, sejam $J = \{k\}$ e $u_i := -\mu(w_i, w_k) =: -v_i$ para $i = 1, 2, \dots, N$. Assumindo que μ satisfaz a desigualdade triangular, prova-se agora a igualdade em (4.10):

$$MK(P, Q) = p_k \sum_{i=1, i \neq k}^N \lambda_i \mu(w_k, w_i).$$

Pela desigualdade triangular, $u_i + v_j = -\mu(w_i, w_k) + \mu(w_j, w_k) \leq \mu(w_i, w_j)$ vale para todo $i, j \in \{1, 2, \dots, N\}$.

Deste modo, pela representação dual (3.9), chega-se ao seguinte resultado:

$$\begin{aligned} MK(P, Q) &\geq \sum_{i=1}^N u_i p_i + \sum_{l \in I \setminus J} v_l q_l \\ &= \sum_{i=1}^N -\mu(w_i, w_k) p_i + \sum_{l \neq k} \mu(w_l, w_k) q_l \\ &= \sum_{i=1, i \neq k}^N \mu(w_i, w_k) (q_i - p_i) \\ &= \mu(w_1, w_k) (q_1 - p_1) + \mu(w_2, w_k) (q_2 - p_2) + \dots + \mu(w_{k-1}, w_k) (q_{k-1} - p_{k-1}) \\ &\quad + \mu(w_{k+1}, w_k) (q_{k+1} - p_{k+1}) + \dots + \mu(w_N, w_k) (q_N - p_N) \\ &= \mu(w_1, w_k) (p_1 + \lambda_1 p_k - p_1) + \dots + \mu(w_N, w_k) (p_N + \lambda_N p_k - p_N) \\ &= p_k \sum_{i=1, i \neq k}^N \lambda_i \mu(w_i, w_k). \end{aligned}$$

Logo,

$$MK(P, Q) \geq \sum_{i=1, i \neq k}^N p_k \lambda_i \mu(w_i, w_k). \quad (4.11)$$

Portanto, por (4.10), o limite superior para $MK(P, Q)$ está provado. A igualdade no caso $\#J = 1$ e μ satisfazendo a desigualdade triangular provada por (4.10) e (4.11).

A desigualdade (4.10) é em geral estrita se J contém mais de um elemento ou se a função μ não satisfaz a desigualdade triangular. É importante observar que esta última condição para μ não é necessária para a validade do Teorema 3 (redistribuição ótima).

Agora que já se sabe como redistribuir os cenários preservados, falta discutir como proceder para escolher os índices do conjunto J (com cardinalidade fixada $N - N_{red}$) de maneira ótima. Este será o próximo assunto a ser tratado.

4.2 – Escolha Ótima de Cenários

Esta seção apresenta técnicas heurísticas para resolver o problema combinatório de ROC. Os algoritmos correspondentes são descritos em detalhes no Capítulo 5.

Conhecida a cardinalidade $N - N_{red}$ do conjunto J^* , quer-se escolher índices de cenários j_1, j_2, \dots, j_{N-n} de forma que a distância entre os valores $V(P)$ e $V(Q)$, definidos em (3.19), seja mínima. Para isso, aplica-se a redistribuição ótima do Teorema 3, e selecionam-se os índices que apresentam o menor valor de $c(J)$ segundo dado em (4.6).

Considere agora dois exemplos extremos do problema de avaliar $c(J)$. Isto é, o caso $\#J = 1$ e o caso $\#J = N - 1$.

Exemplo 2 *Descarte de um único cenário.*

Se $\#J = 1$, o problema de redução ótima (4.6) tem a forma

$$\min_{j \in \{1, 2, \dots, N\}} p_j \min_{l \neq j} \mu(w_j, w_l). \quad (4.12)$$

Se o mínimo é obtido em $j_1^* \in \{1, 2, \dots, N\}$, o cenário $w_{j_1^*}$ é descartado e, pela regra de distribuição ótima, $q_l = p_l$ para cada $l \notin \{j_1^*, l^*\}$, para l^* o índice do cenário mais próximo de $w_{j_1^*}$ segundo alguma função μ , isto é, $l^* = j(j_1^*) = \arg \min_{l \neq j_1^*} \mu(w_{j_1^*}, w_l)$. Neste caso, $q_{l^*} = p_{l^*} + p_{j_1^*}$.

Para um exemplo numérico, considere os cenários da Figura 4.2.

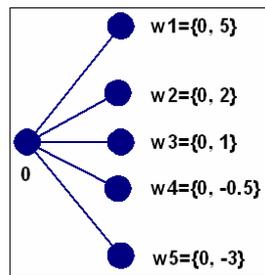


Figura 4.2: Cenários.

Considere também, a matriz μ e o vetor de probabilidades P seguintes:

$$\mu = \begin{bmatrix} 0 & 3 & 4 & 5.5 & 8 \\ 3 & 0 & 1 & 2.5 & 5 \\ 4 & 1 & 0 & 1.5 & 4 \\ 5.5 & 2.5 & 1.5 & 0 & 2.5 \\ 8 & 5 & 4 & 2.5 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} P(w_1) = p_1 = \frac{7}{30} \\ P(w_2) = p_2 = \frac{1}{6} \\ P(w_3) = p_3 = \frac{1}{5} \\ P(w_4) = p_4 = \frac{1}{5} \\ P(w_5) = p_5 = \frac{1}{5} \end{bmatrix}.$$

Cada elemento μ_{ij} da matriz μ representa a distância do cenário w_i ao cenário w_j , ou seja, $\mu(w_i, w_j)$. A função μ escolhida é o valor absoluto $\mu(w_i, w_j) = |w_i - w_j|$.

Para determinar o cenário a ser descartado, analisa-se a degradação da árvore

causada pelo descarte de cada um dos cenários diferentes de $w_{j_1}^*$. O cenário que oferecer menor degradação, segundo calculada por (4.6), será eliminado. Tenta-se agora detalhar melhor tal procedimento.

Faz-se $J = \{1\}$, e por (4.6) tem-se:

$$c(J) = p_1 \min_{l \neq 1} \mu(w_l, w_l) = p_1 \mu_{12} = \frac{7}{30} 3 = \frac{7}{10} \Rightarrow J_1 = j(1) = \{2\}.$$

Agora, para $J = \{2\}$,

$$c(J) = p_2 \min_{l \neq 2} \mu(w_l, w_l) = p_2 \mu_{23} = \frac{1}{6} 1 = \frac{1}{6} \Rightarrow J_2 = j(2) = \{3\}.$$

Para $J = \{3\}$,

$$c(J) = p_3 \min_{l \neq 3} \mu(w_l, w_l) = p_3 \mu_{32} = \frac{1}{5} 1 = \frac{1}{5} \Rightarrow J_3 = j(3) = \{2\}.$$

Para $J = \{4\}$,

$$c(J) = p_4 \min_{l \neq 4} \mu(w_l, w_l) = p_4 \mu_{43} = \frac{1}{5} \frac{3}{2} = \frac{3}{10} \Rightarrow J_4 = j(4) = \{3\}.$$

E, por último, $J = \{5\}$

$$D_{\{5\}} = p_5 \min_{l \neq 5} \mu(w_l, w_l) = p_5 \mu_{54} = \frac{1}{5} \frac{5}{2} = \frac{1}{2} \Rightarrow J_5 = j(5) = \{4\}.$$

Logo, o cenário w_2 é descartado, pois o conjunto $J = \{2\}$ oferece menor valor para $c(J)$. As novas probabilidades são dadas por:

$$\left[\begin{array}{l} q_1 = \frac{7}{30} = p_1 \\ q_2 = 0 \\ q_3 = \frac{11}{30} = p_3 + p_2 \\ q_4 = \frac{1}{5} = p_4 \\ q_5 = \frac{1}{5} = p_5 \end{array} \right]$$

Logo $Q^* = \left\{ \frac{7}{30}, \frac{11}{30}, \frac{1}{5}, \frac{1}{5} \right\}$, e os cenários preservados são mostrados na Figura 4.3 .

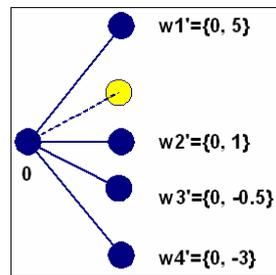


Figura 4.3: Descarte de um cenário.

Exemplo 3 *Preservando um único cenário.*

Se $\#J = N - 1$, a expressão (4.6) tem a seguinte forma:

$$\min_{u \in \{1, 2, \dots, N\}} \sum_{i=1, i \neq u}^N p_i \mu(w_i, w_u). \quad (4.13)$$

Suponha que o mínimo é obtido em $u^* \in \{1, 2, \dots, N\}$, então o cenário w_{u^*} é mantido e, pela regra de redistribuição de probabilidades (4.7), $q_{u^*} = p_{u^*} + \sum_{i \neq u^*} p_i = 1$.

Considerando novamente a situação numérica do Exemplo 2, tem-se a seguinte situação:

Colocando $J = \{2, 3, 4, 5\}$,

$$\begin{aligned} c(J) &= p_2 \mu_{21} + p_3 \mu_{31} + p_4 \mu_{41} + p_5 \mu_{51} \\ &= \frac{1}{6} 3 + \frac{1}{5} 4 + \frac{1}{5} \frac{11}{2} + \frac{1}{5} 8 = 4. \end{aligned}$$

Agora, sendo $J = \{1, 3, 4, 5\}$,

$$\begin{aligned} c(J) &= p_1 \mu_{12} + p_3 \mu_{32} + p_4 \mu_{42} + p_5 \mu_{52} \\ &= \frac{7}{30} 3 + \frac{1}{5} 1 + \frac{1}{5} \frac{5}{2} + \frac{1}{5} 5 = 2.4. \end{aligned}$$

Se $J = \{1, 2, 4, 5\}$,

$$\begin{aligned} c(J) &= p_1 \mu_{13} + p_2 \mu_{23} + p_4 \mu_{43} + p_5 \mu_{53} \\ &= \frac{7}{30} 4 + \frac{1}{6} 1 + \frac{1}{5} \frac{3}{2} + \frac{1}{5} 4 = 2.2. \end{aligned}$$

Se $J = \{1, 2, 3, 5\}$,

$$\begin{aligned} c(J) &= p_1 \mu_{14} + p_2 \mu_{24} + p_3 \mu_{34} + p_5 \mu_{54} \\ &= \frac{7}{30} \frac{11}{2} + \frac{1}{6} \frac{5}{2} + \frac{1}{5} \frac{3}{2} + \frac{1}{5} \frac{5}{2} = 2.5. \end{aligned}$$

Finalmente, se $J = \{1,2,3,4\}$,

$$\begin{aligned} c(J) &= p_1\mu_{13} + p_2\mu_{23} + p_3\mu_{35} + p_4\mu_{45} \\ &= \frac{7}{30}8 + \frac{1}{6}5 + \frac{1}{5}4 + \frac{1}{5}\frac{5}{2} = 4. \end{aligned}$$

Logo, o cenário w_3 é preservado, pois o conjunto $J = \{1,2,4,5\}$ oferece menor valor para $c(J)$. O cenário preservado, que neste caso possui probabilidade igual a 1, é mostrado na Figura 4.4.

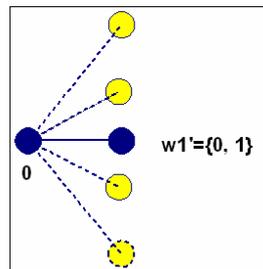


Figura 4.4: Descarte de $N-I$ cenários.

Pode-se, ao invés de determinar antecipadamente a cardinalidade de J^* , usar uma tolerância de dispersão para a ROC. Assim, com o uso de (4.6) e (4.7), é deduzida a chamada *estratégia de redução máxima (erm)* para determinar o conjunto J^* e a distribuição de probabilidades Q^* , de todos cenários preservados ($w_l \in I \setminus J^*$). Esta estratégia calcula o conjunto de índices de cenários descartados com cardinalidade máxima tal que $c(J^*)$ fique menor ou igual a uma tolerância. Em outras palavras, é exigido que $V(Q^*)$ permaneça próximo a $V(P)$ com uma certa exatidão.

A estratégia de redução máxima se resume em dois importantes passos:

1) Determinar o conjunto J^* de índices de cenários, com cardinalidade máxima, tal que

$$\sum_{j \in J} p_j \min_{l \in I \setminus J^*} \mu(w_j, w_l) \leq \varepsilon, \text{ para } \varepsilon > 0 \text{ dado.} \quad (4.14)$$

2) Aplicar a regra de redistribuição (4.7) para cada cenário $w_l, l \in I \setminus J^*$.

O próximo resultado estabelece limites inferior e superior para o valor ótimo de (4.6), que correspondem às extensões recursivas das estratégias (4.12) e (4.13).

Além disso, como mostra [6], o limite inferior é obtido sob hipóteses adicionais.

Teorema 5 Limites. Para o problema de ROC (4.5) e para $i = 1, \dots, N - N_{red}$, $j = 1, \dots, N_{red}$, sejam os cenários l_i e s_j escolhidos recursivamente como as respectivas soluções dos seguintes problemas de minimização:

$$\min_{l \in I \setminus \{l_1, \dots, l_{i-1}\}} p_l \min_{j \neq l} \mu(w_l, w_j) \quad (4.15)$$

e

$$\min_{s \notin \{s_1, \dots, s_{j-1}\}} \sum_{i \in \{s_1, \dots, s_{j-1}, s\}} p_i \min_{l \in \{s_1, \dots, s_{j-1}, s\}} \mu(w_l, w_i). \quad (4.16)$$

Para o conjunto $J_s := I \setminus \{s_1, \dots, s_{N_{red}}\}$, as seguintes desigualdades são satisfeitas:

$$\begin{aligned} \sum_{i=1}^{N-N_{red}} p_{l_i} \min_{j \neq l_i} \mu(w_{l_i}, w_j) &\leq \min\{c(J) : J \subset I, \#J = N - N_{red}\} \\ &\leq \sum_{i \in J_s} p_i \min_{j \in \{s_1, \dots, s_{N_{red}}\}} \mu(w_i, w_j). \end{aligned}$$

Além disso, o conjunto de índices $\{l_1, \dots, l_{N-N_{red}}\}$ é uma solução de (4.18) se para cada $i = 1, \dots, N - N_{red}$ o conjunto $\arg \min_{j \in I \setminus \{l_i\}} \mu(w_{l_i}, w_j) \setminus \{l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_{N-n}\}$ é não vazio.

Prova. Para qualquer índice do conjunto $J = \{j_1, \dots, j_{N-N_{red}}\} \subset I$ com $\#J = N - N_{red}$, pelo Teorema 3, vale que

$$\begin{aligned} c(J) &= \sum_{j \in J} p_j \min_{l \in I \setminus J} \mu(w_j, w_l) \\ &= \sum_{i=1}^{N-N_{red}} p_{j_i} \min_{l \in I \setminus \{j_1, \dots, j_{N-n}\}} \mu(w_{j_i}, w_l) \\ &\geq \sum_{i=1}^{N-N_{red}} p_{j_i} \min_{l \in I \setminus \{j_i\}} \mu(w_{j_i}, w_l) \\ &\geq \sum_{i=1}^{N-N_{red}} p_{l_i} \min_{j \in I \setminus \{l_i\}} \mu(w_{l_i}, w_j), \end{aligned}$$

a primeira desigualdade acima provém do fato de que o conjunto viável do problema (4.8) passa a ser menos restrito; a segunda desigualdade segue da definição (4.15) dos índices l_i , para $i = 1, \dots, N - N_{red}$.

Se o conjunto $\arg \min_{j \in I \setminus \{l_i\}} \mu(w_{l_i}, w_j) \setminus \{l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_{N-N_{red}}\}$ é não vazio para cada $i = 1, \dots, N - N_{red}$, então

$$\min_{j \in I \setminus \{l_i\}} \mu(w_{l_i}, w_j) = \min_{j \in I \setminus \{l_1, \dots, l_{N-N_{red}}\}} \mu(w_{l_i}, w_j).$$

Portanto, a estimativa acima pode ser estendida como:

$$\begin{aligned}
 c(J) &\geq \sum_{i=1}^{N-N_{red}} p_{l_i} \min_{j \in I \setminus \{l_1, \dots, l_{N-N_{red}}\}} \mu(w_{l_i}, w_j) \\
 &= \sum_{l \in \{l_1, \dots, l_{N-N_{red}}\}} p_l \min_{j \in I \setminus \{l_1, \dots, l_{N-N_{red}}\}} \mu(w_l, w_j) \\
 &= c(\{l_1, \dots, l_{N-N_{red}}\}).
 \end{aligned}$$

Assim, o conjunto de índices $\{l_1, \dots, l_{N-n}\}$ é solução de (4.8).

Por último, tem-se a estimativa:

$$\min\{c(J) : J \subset I, \#J = N - N_{red}\} \leq c(J_s) = \sum_{j \in J_s} p_j \min_{l \in I \setminus J_s} \mu(w_j, w_l), \quad \text{e assim o}$$

teorema está provado. ■

O Teorema 5 sugere dois diferentes algoritmos heurísticos para a redução ótima de N cenários originais a N_{red} cenários preservados; um algoritmo baseado no limite inferior de $c(J)$, do tipo regressivo, e outro algoritmo que usa o limite superior, do tipo progressivo.

Em resumo, foram apresentadas neste capítulo fórmulas explícitas para a representação da métrica $d_{FM}(P, Q)$ e para a redistribuição ótima de probabilidades. Foi mostrada também a estratégia de redução máxima, que é a principal técnica de ROC. A idéia básica do procedimento dos algoritmos progressivo e regressivo de redução foi apresentada, superficialmente, pelos Exemplos 2 e 3. No próximo capítulo serão propostas e desenvolvidas diferentes variantes destes algoritmos. Estes algoritmos, quando associados com a estratégia de redução máxima, proporcionam o conjunto ótimo de índices de cenários J^* . Serão apresentados também exemplos práticos e figuras ilustrativas para uma melhor compreensão dos procedimentos utilizados.

5 - Algoritmos de Seleção e Algoritmos de Redução de Árvores de Cenários

Este capítulo trata a questão de encontrar o conjunto ótimo de índices de cenários a serem descartados. Segundo mencionado no Capítulo 4, não é possível analisar cada um dos candidatos ao conjunto J^* . Neste sentido, 6 apresenta algoritmos capazes de selecionar índices ótimos que dispensam a análise exaustiva de todas as combinações de cenários.

A partir dos limites para $c(J)$, enunciados e provados pelo Teorema 5, serão formuladas duas técnicas diferentes para a ROC. A primeira técnica determina os $N - N_{red}$ cenários que serão descartados por meio da expressão (4.15), explorando o limite inferior do Teorema 5. A segunda técnica realiza a seleção dos índices a serem preservados por meio da expressão (4.16), a partir do limite superior do Teorema. Para uma melhor compreensão dos procedimentos efetuados pelos algoritmos, serão também apresentados neste capítulo exemplos numéricos e figuras ilustrativas.

Os algoritmos desenvolvidos através dos limites inferior e superior são chamados de redução regressiva e seleção progressiva, respectivamente.

5.1 – Algoritmos para Seleção de Cenários

Parâmetro: N_{red} , o número de cenários a serem preservados, satisfazendo $N_{red} < N$.

Algoritmo 1 Redução Regressiva.

Passo 0: Calcular as distâncias entre pares de cenários:

$$\mu_{kj} = \mu(w_k, w_j), \text{ para } k, j = 1, \dots, N$$

Passo 1: Para $i=1$, faz-se $J^{[0]} = \emptyset$ e seleciona-se, resolvendo iterativamente o problema

$$j_i \in \arg \min_{l \in I \setminus J^{[i-1]}} \sum_{k \in J^{[i-1] \cup l}} p_k \min_{j \in I \setminus J^{[i-1] \cup l}} \mu_{kj}, \quad (5.1)$$

k_1 cenários com $1 \leq k_1 \leq N - N_{red}$, de forma que $J^{[1]} = \{j_1^{(1)}, j_2^{(1)}, \dots, j_{k_1}^{(1)}\}$ seja uma solução de (4.8). Deste modo, $J^{[1]}$ é escolhido usando (4.6), e obtém-se a redistribuição de probabilidades Q_1 dada por (4.7). Faz-se $i = i + 1$, e vai até o Passo i .

Passo i : Se $k_{i-1} < N - N_{red}$, a distribuição Q_{i-1} é reduzida a Q_i , e pelo problema (5.1) seleciona-se $J^{[i]} = \{j_1^{(i)}, j_2^{(i)}, \dots, j_{k_i}^{(i)}\}$, onde $1 < k_i = \#J^{[i]} \leq N - N_{red} - k_{i-1}$, com k_i obtido analogamente a k_1 .

Se $\sum_{j=1}^i k_j = N - N_{red}$, pára-se o algoritmo fazendo $J^* = \bigcup_{j=1}^i J^{[j]}$ e aplicando novamente a regra de redistribuição (4.7) para os cenários w_l , com $l \in I \setminus J^*$.

Se não, faz-se $i = i + 1$, e vai até o Passo i .

Como se pode notar, em cada iteração i acrescentam-se índices no conjunto de trabalho $J^{[i-1]}$. O índice j_i a introduzir é determinado pela solução do problema (5.1). O algoritmo pára quando $\#J = N - N_{red}$, para um N_{red} dado.

No caso em que $k_j = 1$ para $j = 1, \dots, i$ é obtida a chamada Redução Regressiva de cenários Passo à Passo. Se for efetuada a regra de redistribuição na última etapa do algoritmo (ao invés de realizar o processo a cada iteração) tem-se a estratégia de redução chamada Redução Regressiva Simultânea, detalhada a seguir.

Assim como na redução regressiva, $J^{[0]} = \emptyset$ e a cada i -ésima iteração, acrescenta-se um único índice ao conjunto de trabalho J . A seguir os procedimentos realizados por este algoritmo.

Parâmetro(s): N_{red} , o número de cenários a serem preservados satisfazendo $N_{red} < N$ e/ou $\varepsilon > 0$, uma tolerância de dispersão máxima utilizada na *erm*.

Algoritmo 2 Redução Regressiva Simultânea.

Passo 0: Calcular as distâncias entre pares de cenários:

$$\mu_{kj} = \mu(w_k, w_j), \text{ para } k, j = 1, \dots, N$$

Passo 1: Calcular, para $l = 1, \dots, N$,

$$\mu_{ll}^{[1]} = \min_{j \neq l} \mu_{lj} \text{ e } z_l^{[1]} = p_l \mu_{ll}^{[1]}.$$

Escolher $l_i \in \arg \min_{l \in \{1, \dots, N\}} z_l^{[1]}$ e definir $J^{[1]} = \{l_i\}$ (índice a ser descartado).

Passo i : Calcular, para $l \in I \setminus J^{[i-1]}$ e $k \in J^{[i-1]} \cup \{l\}$,

$$\mu_{kl}^{[i]} = \min_{j \in I \setminus J^{[i-1]} \cup \{l\}} \mu_{kj} \text{ e } z_l^{[i]} = \sum_{k \in J^{[i-1]} \cup \{l\}} p_k \mu_{kl}^{[i]}, l \notin J^{[i-1]}.$$

Escolher $l_i \in \arg \min_{l \in I \setminus J^{[i-1]}} z_l^{[i]}$ e definir $J^{[i]} = J^{[i-1]} \cup \{l_i\}$ (índice a ser

descartado).

Se $\sum_{j \in J^{[i]}} p_j \min_{l \in I \setminus J^{[i-1]}} \mu_{jl}^{[i]} > \varepsilon$, e/ou, se $\#J^{[i]} = N - N_{red}$, ir até o passo $N - N_{red}$. Se não, fazer $i = i + 1$ e voltar ao Passo i .

Passo $N - N_{red}$: $J^* = J^{[i]}$, é o conjunto de índices que será descartado.

Finalmente, faz-se o cálculo das probabilidades ótimas para os cenários preservados pela regra (4.7).

O Algoritmo 2 tem a seguinte interpretação: O Passo 1 corresponde ao descarte ótimo de um único cenário. Para $i > 1$, j_i é escolhido de tal forma que $c(J^{[i-1]} \cup j_i) = \min_{j \in I \setminus J^{[i-1]}} c(J^{[i-1]} \cup j)$.

Assim, dado que na $(i-1)$ -ésima iteração o conjunto ótimo a ser descartado é $J^{[i-1]} = \{j_1, j_2, \dots, j_{i-1}\}$, o índice seguinte j_i é escolhido de tal forma que os elementos $\{j_1, j_2, \dots, j_{i-1}, j_i\}$ formam o conjunto ótimo para a i -ésima iteração, ou seja, j_i é obtido por (5.1).

Exemplo 4 Para uma melhor compreensão do procedimento do Algoritmo 2, considere a situação do Exemplo 2, com uma tolerância de dispersão máxima $\varepsilon = 0.5$.

Seja μ a matriz das distâncias de todos pares de cenários ($\mu_{kj} = \mu(w_k, w_j)$), $k, j = 1, 2, 3, 4, 5$, e P as probabilidades dadas, respectivamente por

$$\mu = \begin{bmatrix} 0 & 3 & 4 & 5.5 & 8 \\ 3 & 0 & 1 & 2.5 & 5 \\ 4 & 1 & 0 & 1.5 & 4 \\ 5.5 & 2.5 & 1.5 & 0 & 2.5 \\ 8 & 5 & 4 & 2.5 & 0 \end{bmatrix} \quad \begin{bmatrix} P(w_1) = p_1 = \frac{7}{30} \\ P(w_2) = p_2 = \frac{1}{6} \\ P(w_3) = p_3 = \frac{1}{5} \\ P(w_4) = p_4 = \frac{1}{5} \\ P(w_5) = p_5 = \frac{1}{5} \end{bmatrix}.$$

Foi verificado no Capítulo 4 (Exemplo 2) que na primeira iteração o melhor cenário a ser descartado é w_2 , ou seja, $J^{[1]} = \{2\}$. Mas como $c(J^{[1]}) = \frac{1}{6} < \varepsilon$, o algoritmo deve continuar selecionando índices para descartar. A escolha do próximo índice é feita segundo explicado a seguir.

Faz-se $J^{[2]} = \{2, 1\}$ e calcula-se $c(J^{[2]})$.

$$c(J^{[2]}) = \sum_{j \in \{2,1\}} p_j \min_{l \in I \setminus \{2,1\}} \mu(w_j, w_l) = p_1 \mu_{13} + p_2 \mu_{23} = \frac{7}{30} 4 + \frac{1}{6} 1 = \frac{33}{30}.$$

Para $J^{[2]} = \{2, 3\}$, tem-se:

$$c(J^{[2]}) = \sum_{j \in \{2,3\}} p_j \min_{l \in I \setminus \{2,3\}} \mu(w_j, w_l) = p_2 \mu_{24} + p_3 \mu_{34} = \frac{1}{6} 5 + \frac{1}{5} 3 = \frac{43}{60}.$$

Para $J^{[2]} = \{2, 4\}$,

$$c(J^{[2]}) = \sum_{j \in \{2,4\}} p_j \min_{l \in I \setminus \{2,4\}} \mu(w_j, w_l) = p_2 \mu_{23} + p_4 \mu_{43} = \frac{1}{6} 1 + \frac{1}{5} 3 = \frac{14}{30}.$$

Por último, para $J^{[2]} = \{2, 5\}$,

$$c(J^{[2]}) = \sum_{j \in \{2,5\}} p_j \min_{l \in I \setminus \{2,5\}} \mu(w_j, w_l) = p_2 \mu_{23} + p_5 \mu_{54} = \frac{1}{6} + \frac{1}{5} \frac{5}{2} = \frac{2}{3}.$$

Logo, os índices $\{2,4\}$ proporcionaram o menor valor de $c(J^{[2]})$. Portanto, na segunda iteração o conjunto ótimo é $J^{[2]} = \{2,4\}$.

Como $c(J^{[2]}) = 0.4667 < 0.5$, o algoritmo continua a selecionar índices.

Faz-se $J^{[3]} = \{2,4,1\}$ e calcula-se

$$c(J^{[3]}) = p_2 \mu_{23} + p_4 \mu_{43} + p_1 \mu_{13} = \frac{1}{6} + \frac{1}{5} \frac{3}{2} + \frac{7}{30} \frac{3}{2} = \frac{49}{60}.$$

Para $J^{[3]} = \{2,4,3\}$, tem-se

$$c(J^{[3]}) = p_2 \mu_{21} + p_4 \mu_{45} + p_3 \mu_{31} = \frac{1}{6} + \frac{1}{5} \frac{5}{2} + \frac{1}{5} 4 = \frac{9}{5}.$$

Por último, para $J^{[3]} = \{2,4,5\}$,

$$c(J^{[3]}) = p_2 \mu_{23} + p_4 \mu_{43} + p_5 \mu_{53} = \frac{1}{6} + \frac{1}{5} \frac{1}{2} + \frac{1}{5} 4 = \frac{16}{15}.$$

Portanto, na terceira iteração o conjunto ótimo é $J^{[3]} = \{2,4,1\}$, pois proporcionou $c(J^{[3]}) = \frac{49}{60} = 0.8167$. Mas como se nota, a tolerância ε foi ultrapassada ($c(J^{[3]}) > 0.5$).

Nestas condições, este último cenário não deve ser descartado. Então, o conjunto de índices a ser descartado resulta $J^* = J^{[2]} = \{2,4\}$.

A Figura 5.1 mostra a evolução do processo de seleção dos cenários a serem descartados segundo o Algoritmo 2.

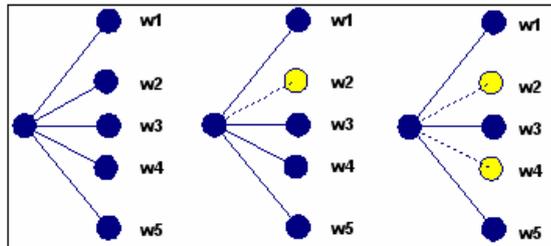


Figura 5.1: Evolução do processo de seleção de cenários (regressiva).

Passa-se agora, utilizando a regra de distribuição (4.7), ao cálculo das probabilidades ótimas para os cenários preservados. Verifica-se que dentre os cenários preservados, o cenário 3 está mais próximo do cenário descartado 2, isto é, $\mu_{23} = \mu(w_2, w_3) \leq \mu(w_2, w_l)$ para todo $l \in I \setminus J$. Do mesmo modo, o cenário 3 está mais próximo do cenário descartado 4 ($\mu_{43} = \mu(w_4, w_3) \leq \mu(w_4, w_l)$ para todo $l \in I \setminus J$). Assim,

$$3 = l(2) \in \arg \min_{l \in J} \mu(w_2, w_l)$$

$$3 = l(4) \in \arg \min_{l \in J} \mu(w_4, w_l)$$

Logo, $J_3 = \{2,4\}$ e $J_1 = J_5 = \emptyset$, pois os cenários 1 e 5 não são os mais próximos dos cenários descartados.

Fazendo $q_l = p_l + \sum_{j \in J_l} p_j$ para cada $l \in I \setminus J$, tem-se:

$$q_1 = p_1 = \frac{7}{30}$$

$$q_2 = 0$$

$$q_3 = p_3 + p_2 + p_4 = \frac{1}{5} + \frac{1}{6} + \frac{1}{5} = \frac{17}{30}$$

$$q_4 = 0$$

$$q_5 = p_5 = \frac{1}{5}$$

Logo, a distribuição de probabilidades $Q^* = \{q_1, q_2, q_3\}$ é igual a $\{\frac{7}{30}, \frac{17}{30}, \frac{1}{5}\}$, e os cenários preservados são mostrados na Figura 5.2.

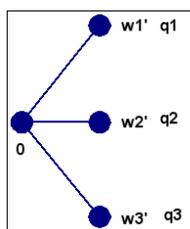


Figura 5.2: Cenários preservados.

Em relação ao Algoritmo 2, é importante notar que, a medida que a cardinalidade de J aumenta, o esforço computacional aumenta de forma combinatória. Quando N é grande e desejam-se descartar muitos cenários, o Algoritmo 2 pode envolver um trabalho computacional extremamente elevado. Para contornar esta dificuldade, é utilizada outra estratégia, de redução progressiva, que será apresentada a seguir.

Como nos algoritmos anteriores, o algoritmo de seleção progressiva também se baseia na estratégia de redução máxima (*erm*). No entanto, a seleção de índices é realizada de forma oposta: a princípio todos cenários são descartados ($J^{[0]} = I$).

A cada iteração i , para $i = 1, \dots, N_{red}$, é escolhido um índice u_i de cenário a ser preservado, ou seja, $J^{[k]} = I \setminus \{u_1, \dots, u_k\}$. Os índices u_i são determinados de forma recursiva, com base no limite superior do Teorema 5, mais especificamente como em (4.16).

Parâmetro(s): N_{red} , o número de cenários a serem preservados satisfazendo $N_{red} < N$ e/ou $\varepsilon > 0$, uma tolerância de dispersão máxima utilizada na *erm*.

Algoritmo 3 Seleção Progressiva

Passo 0: Faz-se $J^{[0]} = \{1, 2, \dots, N\}$ e obtém-se as distâncias entre pares de cenários:

$$\mu_{kj} = \mu(w_k, w_j), \text{ para } k \text{ e } j = 1, \dots, N.$$

Passo i: Calcular

$$u_i \in \arg \min_{u \in J^{[i-1]}} \left\{ c(J^{[i-1]} \setminus \{u\}) = \sum_{j \in J^{[i-1]} \setminus \{u\}} p_j \min_{l \in J^{[i-1]} \setminus \{u\}} \mu(w_j, w_l) \right\}, \quad (5.2)$$

e definir: $J^{[i]} = J^{[i-1]} \setminus \{u_i\}$.

Se $c(J^{[i]})$ for menor ou igual a uma tolerância de dispersão prefixada ($c(J^{[i]}) \leq \varepsilon$), e/ou, se $\#J^{[i]} = N - N_{red}$, ir até o Passo $N - N_{red}$. Senão, fazer $i = i + 1$ e voltar ao Passo i.

Passo $N - N_{red}$: O conjunto de índices a ser descartados é $J^* = J^{[i]}$. Fazer o cálculo das probabilidades ótimas, usando a regra (4.7) para os cenários preservados.

Exemplo 5 Considerando a mesma situação do Exemplo 2, procura-se descartar o maior número possível de cenários dada uma tolerância de dispersão $\varepsilon = 0.5$. Para esta redução será utilizado o Algoritmo 3.

Faz-se $J^{[0]} = \{1, 2, 3, 4, 5\}$, ou seja, a priori todos os cenários serão descartados. No primeiro passo procura-se um índice u tal que w_u seja preservado. Para selecionar u , testam-se todos os cenários fazendo $J^{[1]} = \{1, 2, 3, 4, 5\} \setminus \{i\}$ para $i = 1, 2, 3, 4, 5$ e define-se $u := i^* \in \operatorname{argmin} c(J^{[1]})$.

Considerando o exemplo numérico tem-se:

Passo 1:

Para $J^{[1]} = \{2, 3, 4, 5\}$,

$$\begin{aligned} c(\{2, 3, 4, 5\}) &= p_2 \mu_{21} + p_3 \mu_{31} + p_4 \mu_{41} + p_5 \mu_{51} \\ &= \frac{1}{6} 3 + \frac{1}{5} 4 + \frac{1}{5} 5 + \frac{1}{5} 8 \\ &= 3.9; \end{aligned}$$

Para $J^{[1]} = \{1, 3, 4, 5\}$,

$$\begin{aligned} c(\{1, 3, 4, 5\}) &= p_1 \mu_{12} + p_3 \mu_{32} + p_4 \mu_{42} + p_5 \mu_{52} \\ &= \frac{7}{30} 3 + \frac{1}{5} 1 + \frac{1}{5} 2 + \frac{1}{5} 5 \\ &= 2.4; \end{aligned}$$

Para $J^{[1]} = \{1, 2, 4, 5\}$,

$$\begin{aligned}
 c(\{1,2,4,5\}) &= p_1\mu_{13} + p_2\mu_{23} + p_4\mu_{43} + p_5\mu_{53} \\
 &= \frac{7}{30}4 + \frac{1}{6}1 + \frac{1}{5}\frac{3}{2} + \frac{1}{5}4 \\
 &= 2.2;
 \end{aligned}$$

Para $J^{[1]} = \{1,2,3,5\}$,

$$\begin{aligned}
 c(\{1,2,3,5\}) &= p_1\mu_{14} + p_2\mu_{24} + p_3\mu_{34} + p_5\mu_{54} \\
 &= \frac{7}{30}\frac{11}{2} + \frac{1}{6}\frac{5}{2} + \frac{1}{5}\frac{3}{2} + \frac{1}{5}\frac{5}{2} \\
 &= 2.5
 \end{aligned}$$

Para $J^{[1]} = \{1,2,3,4\}$,

$$\begin{aligned}
 c(\{1,2,3,4\}) &= p_1\mu(1,5) + p_2\mu(2,5) + p_3\mu(3,5) + p_4\mu(4,5) \\
 &= \frac{7}{30}8 + \frac{1}{6}5 + \frac{1}{5}4 + \frac{1}{5}\frac{5}{2} \\
 &= 4
 \end{aligned}$$

Logo, $3 \in \operatorname{argmin}c(J^{[0]} \setminus \{i\})$ para $i = 1,2,3,4,5$. Portanto $J^{[1]} = \{1,2,4,5\}$ e o cenário w_3 será preservado.

Como $c(J^{[1]}) = 2.2 > \varepsilon$, o algoritmo continua com a seleção de índices.

Passo 2:

Para $J^{[2]} = \{2,4,5\}$,

$$\begin{aligned}
 c(\{2,4,5\}) &= p_2\mu_{23} + p_4\mu_{43} + p_5\mu_{53} \\
 &= \frac{1}{6}1 + \frac{1}{5}\frac{3}{2} + \frac{1}{5}4 \\
 &= 1.2667;
 \end{aligned}$$

Para $J^{[2]} = \{1,4,5\}$,

$$\begin{aligned}
 c(\{1,4,5\}) &= p_1\mu_{12} + p_4\mu_{43} + p_5\mu_{53} \\
 &= \frac{7}{30}3 + \frac{1}{5}\frac{3}{2} + \frac{1}{5}4 \\
 &= 1.8;
 \end{aligned}$$

Para $J^{[2]} = \{1,2,5\}$,

$$\begin{aligned}
 c(\{1,2,5\}) &= p_1\mu_{13} + p_2\mu_{23} + p_5\mu_{54} \\
 &= \frac{7}{30}4 + \frac{1}{6}1 + \frac{1}{5}\frac{5}{2} \\
 &= 1.6;
 \end{aligned}$$

Para $J^{[2]} = \{1,2,4\}$,

$$\begin{aligned}
 c(\{1,2,4\}) &= p_1\mu_{13} + p_2\mu_{23} + p_4\mu_{53} \\
 &= \frac{7}{30}4 + \frac{1}{6}1 + \frac{1}{5}\frac{3}{2} \\
 &= 1.4;
 \end{aligned}$$

Assim, $1 \in \operatorname{argmin}c(J^{[1]} \setminus \{i\})$ para $i = 1,2,4,5$. Portanto $J^{[2]} = \{2,4,5\}$ e o cenário

w_1 também será preservado.

$c(J^{[2]}) = 1.2667 > \varepsilon$, o algoritmo continua com a seleção de índices.

Passo 3:

Para $J^{[3]} = \{4,5\}$,

$$\begin{aligned} c(\{4,5\}) &= p_4\mu_{43} + p_5\mu_{53} \\ &= \frac{1}{5} \frac{3}{2} + \frac{1}{5} 4 \\ &= 1.1; \end{aligned}$$

Para $J^{[3]} = \{2,4\}$,

$$\begin{aligned} c(\{2,4\}) &= p_2\mu_{23} + p_4\mu_{43} \\ &= \frac{1}{6} 1 + \frac{1}{5} \frac{3}{2} \\ &= 0.4667; \end{aligned}$$

Para $J^{[3]} = \{2,5\}$,

$$\begin{aligned} c(\{2,5\}) &= p_2\mu_{23} + p_5\mu_{54} \\ &= \frac{1}{6} 1 + \frac{1}{5} \frac{5}{2} \\ &= 0.6667 \end{aligned}$$

Assim, $5 \in \operatorname{argmin} c(J^{[2]} \setminus \{i\})$ para $i = 2,4,5$. Portanto $J^{[3]} = \{2,4\}$ e o cenário w_5 também será preservado.

Como $c(J^{[3]}) = 0.4667 < \varepsilon = 0.5$, o algoritmo pára e tem-se $J^* = J^{[3]} = \{2,4\}$.

Realizando a redistribuição pela regra (4.7), passa-se a ter a seguinte distribuição de probabilidades para os cenários preservados como no exemplo 4:

$$Q^* = \begin{bmatrix} q_1 = \frac{7}{30} \\ q_2 = \frac{17}{30} \\ q_3 = \frac{1}{5} \end{bmatrix}$$

A Figura 5.3 mostra a evolução dos índices a serem preservados.

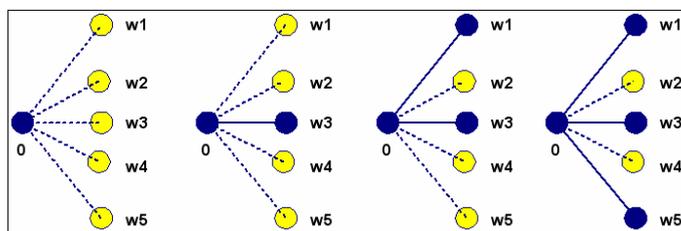


Figura 5.3: Evolução do processo de seleção de cenários (progressivo).

Como se pode notar, os resultados obtidos pelos Algoritmos 3 e 2 são idênticos. A

única diferença entre ambos procedimentos é apenas o referencial de escolha iterativa: um seleciona índices a serem descartados (algoritmo regressivo) e o outro seleciona índices a serem preservados (algoritmo progressivo). Dada uma tolerância igual a 0.5, ambos algoritmos descartaram os cenários w_2 e w_4 . Há, no entanto, situações em que os procedimentos não proporcionam exatamente o mesmo conjunto de índices J . Isto se deve ao fato de que descartar o cenário k seja equivalente, em termos da função $c(J)$, a descartar o cenário j , para $k \neq j$.¹²

Quando $N_{red} < N - N_{red}$, ou seja, o número de cenários preservados é menor que o número de cenários descartados, é conveniente usar a estratégia progressiva (já que o número de escolhas a serem realizadas é menor). Por outro lado, quando $N_{red} > N - N_{red}$, a estratégia regressiva se torna computacionalmente mais atrativa. O artigo [10] apresenta resultados mais exatos a respeito da complexidade computacional de cada estratégia.

Assim como o Algoritmo 2 é uma variante da redução regressiva, o Algoritmo 4 a seguir, é uma variante da seleção progressiva, chamado de Seleção Progressiva Rápida. Sua principal vantagem é diminuir o esforço computacional exigido pelo Algoritmo 3.

Parâmetro(s): N_{red} , o número de cenários a serem preservados satisfazendo $N_{red} < N$ e/ou $\varepsilon > 0$, uma tolerância de dispersão máxima utilizada na *erm*.

Algoritmo 4 Seleção Progressiva Rápida

Passo 0: Faz-se $J^{[0]} = \{1, 2, \dots, N\}$ e obtém-se as distâncias entre pares de cenários:

$$\mu_{ku}^{[1]} = \mu(w_k, w_u), \text{ para } k \text{ e } u = 1, \dots, N.$$

Passo 1: Calcular

$$z_u^{[1]} = \sum_{k=1, k \neq u}^N p_k \mu_{ku}^{[1]}, \quad u = 1, \dots, N.$$

Escolher $u_1 \in \arg \min_{u \in \{1, \dots, N\}} z_u^{[1]}$ e fazer $J^{[1]} = \{1, \dots, N\} \setminus \{u_1\}$.

Passo i : Calcular

$$\mu_{ku}^{[i]} = \min\{\mu_{ku}^{[i-1]}, \mu_{ku_{i-1}}^{[i-1]}\}, \quad k, u \in J^{[i-1]}, \text{ e}$$

$$z_\mu^{[i]} = \sum_{k \in J^{[i-1]} \setminus \{u\}} p_k \mu_{k\mu}^{[i]}, \quad u \in J^{[i-1]}$$

Escolher $u_i \in \arg \min_{u \in J^{[i-1]}} z_u^{[i]}$ e fazer $J^{[i]} = J^{[i-1]} \setminus u_i$.

Se $z_{u_i}^{[i]} \leq \varepsilon$, ir até o passo $N - N_{red}$, caso contrário, fazer $i = i + 1$ e voltar ao Passo i .

¹² Na ocorrência de tal situação, recomenda-se descartar dentre os dois cenários, aquele que possui maior norma (se $\|w_j\| > \|w_k\|$, descarta-se w_k). Com isto, tenta-se preservar o cenário que está mais próximo a um dos extremos da árvore.

Passo $N - N_{red}$: $J^* = J^{[i]}$ é o conjunto de índices a ser descartados.

Finalmente, faz-se o cálculo das probabilidades ótimas pela regra (4.7) para os cenários preservados.

Poderia se pensar que, pelas definições de $u^{[i]}$ e $z_u^{[i]}$, os algoritmos 3 e 4 são bastantes distintos; mas na verdade, a diferença está apenas na redução de operações básicas requeridas pelo Algoritmo 4, quando comparado ao Algoritmo 3. De fato, as expressões $z_{u_i}^{[i]}$ e $c(J^{[i]})$ proporcionam o mesmo resultado segundo enunciado no próximo teorema.

Teorema 6 O conjunto $I \setminus J = \{u_1, \dots, u_{N_{red}}\}$ determinado pelo Algoritmo 4 é uma solução do algoritmo da seleção progressiva. Mais precisamente, para cada $i = 1, \dots, N_{red}$, o índice u_i satisfaz a condição (5.2) e $z_{u_i}^{[i]} = c(J^{[i]})$, com $c(J^{[i]})$ definido em (4.6).

Prova. Para $i = 1$, tem-se $J^{[i-1]} = J^{[0]} = \{1, 2, \dots, N\}$ e

$$z_{u_1}^{[1]} = \sum_{k \neq u_1} p_k \mu_{ku}^{[1]} = \sum_{k \neq u_1} p_k \min \mu(w_k, w_{u_1}) = \sum_{k \neq u_1} p_k \min_{l \in I \setminus J^{[0]} \setminus \{u_1\}} \mu(w_k, w_l).$$

Seja $u_1 \in \arg \min_{u \in J^{[0]}} z_u^{[1]}$. O conjunto $J^{[1]}$ é dado por $J^{[1]} = J^{[0]} \setminus \{u_1\}$, logo

$$z_{u_1}^{[1]} = \sum_{k \neq u_1} p_k \min_{l \in I \setminus J^{[0]} \setminus \{u_1\}} \mu(w_k, w_l) = \sum_{k \in J^{[1]}} \min_{l \in I \setminus J^{[1]}} \mu(w_k, w_l) = c(J^{[1]}).$$

Considerando agora $i = 2, \dots, N_{red}$, segue-se da aplicação recursiva da definição de $\mu_{ku}^{[i-j]}$, para $j = 0, \dots, i$, que:

$$\begin{aligned} z_{u_i}^{[i]} &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \mu_{ju}^{[i]} \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min(\mu_{ju}^{[i-1]}, \mu_{ju_{i-1}}^{[i-1]}) \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min\{\min(\mu_{ju}^{[i-2]}, \mu_{ju_{i-2}}^{[i-2]}), \min(\mu_{ju_{i-1}}^{[i-2]}, \mu_{ju_{i-2}}^{[i-2]})\} \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min\{\mu_{ju}^{[i-2]}, \mu_{ju_{i-1}}^{[i-2]}, \mu_{ju_{i-2}}^{[i-2]}\} \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min\{\min(\mu_{ju}^{[i-3]}, \mu_{ju_{i-3}}^{[i-3]}), \min(\mu_{ju_{i-1}}^{[i-3]}, \mu_{ju_{i-3}}^{[i-3]}), \min(\mu_{ju_{i-2}}^{[i-3]}, \mu_{ju_{i-3}}^{[i-3]})\} \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min\{\mu_{ju}^{[i-3]}, \mu_{ju_{i-1}}^{[i-3]}, \mu_{ju_{i-2}}^{[i-3]}, \mu_{ju_{i-3}}^{[i-3]}\} \\ &\vdots \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min\{\mu_{ju}, \mu_{ju_{i-1}}, \mu_{ju_{i-2}}, \dots, \mu_{ju_1}\} \\ &= \sum_{j \in J^{[i-1]} \setminus \{u_i\}} p_j \min_{l \in I \setminus J^{[i-1]} \setminus \{u_i\}} \mu_{jl}. \end{aligned}$$

Se $u_i \in \arg \min z_{u_i}^{[i]}$, tem-se que $J^{[i]} = J^{[i-1]} \setminus \{u_i\}$ e, para $i = 1, \dots, N_{red}$,

$$\begin{aligned} z, \mu_i^{[i]} &= \sum_{j \in J^{[i-1]} \setminus \{u\}} p_j \min_{l \in I \setminus J^{[i-1]} \setminus \{u\}} \mu_{jl} \\ &= \sum_{j \in J^{[i]} \setminus \{u\}} p_j \min_{l \in I \setminus J^{[i]}} \mu_{jl} \\ &= c(J^{[i]}), \end{aligned}$$

como se queria provar. ■

Este teorema comprova a equivalência dos Algoritmos 3 e 4. A vantagem de se usar o Algoritmo 4, ao invés do Algoritmo 3, é a conseguinte redução do esforço computacional, pois a cada passo i do algoritmo de Seleção Progressiva Rápida, é calculado apenas o mínimo entre dois valores ($\mu_{ku}^{[i]} = \min\{\mu_{ku}^{[i-1]}, \mu_{ku_{-1}}^{[i-1]}\}$) para cada $k, u \in J^{[i-1]}$. Já no algoritmo de Seleção Progressiva, é calculada a mínima distância entre cada cenário $j \in J$ e todos os cenários $l \in I \setminus J$ ($\min_{l \in I \setminus J} \mu_{jl}$). A desvantagem de se utilizar o Algoritmo 4 é a necessidade de trabalhar com a matriz de distâncias completa¹³. Para os demais algoritmos de seleção de cenários, basta considerar "somente" a parte triangular superior, ou triangular inferior da matriz de distâncias, que é simétrica.

Foram apresentadas nesta seção as principais estratégias de seleção dos índices para a redução de cenários. Suas variantes, Redução Regressiva Simultânea e Seleção Progressiva Rápida, fazem o uso da *erm*. Estes algoritmos são as principais ferramentas no processo de redução de cenários.

No próximo capítulo serão apresentados resultados de redução, utilizando os Algoritmos 2, 3 e 4 para problemas de grande porte. Mas antes disso, serão apresentados algoritmos que trabalham mais especificamente com cenários estruturados em formato árvore.

5.2 – Algoritmos de Redução de Árvores de Cenários

Uma árvore de cenários é uma forma abstrata de se representar e mostrar o comportamento de incertezas ao longo do tempo. Um simples exemplo de cenários no formato árvore foi apresentado na Figura 2.1 do Capítulo 2.

Uma árvore de cenários pode ser representada de forma matricial. Considere a Figura 5.4.

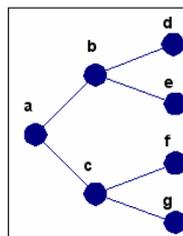


Figura 5.4: Árvore de cenários.

¹³ Dependendo do número N de cenários a matriz pode ser de difícil armazenamento computacional.

Matricialmente, pode-se escrever esta árvore da seguinte forma:

$$\text{Árvore} = \begin{bmatrix} a & b & d \\ a & b & e \\ a & c & f \\ a & c & g \end{bmatrix} \text{ (árvore de cenários com 2 períodos e 4 cenários).}$$

Neste sentido, cada linha da matriz, $w_1 = [a, b, d]$; $w_2 = [a, b, e]$; $w_3 = [a, c, f]$ e $w_4 = [a, c, g]$, corresponde a um cenário.

Serão apresentados a seguir métodos para redução de árvores de cenários.

5.2.1 – Redução e Agregação de Cenários

A redução da árvore de cenários é baseada em reduções sucessivas ao longo dos estágios de tempo t . Assumindo que cada cenário w_i possui um número finito de nós, e considerando T o número de períodos do horizonte de planejamento, w_i é um vetor (pode ser uma matriz, segundo seja o número de variáveis incertas que represente) pertencente a \mathbb{R}^T .

O cenário w_i pode ser escrito na seguinte forma vetorial:

$w_i = \{w_i^t\}_{t=1}^T = (w_i^1, w_i^2, \dots, w_i^T)^{14}$. Assim, considerando p_i^t como a probabilidade de alcançar o nó (i, t) que compõe o cenário w_i no instante de tempo t , a probabilidade associada à ocorrência do cenário w_i é dada pelo produto $p_i = \prod_{t=1}^T p_i^t$.

Além disso, é assumido que todos cenários coincidem no instante de tempo $t = 0$, isto é, $w_1^0 = w_2^0 = w_3^0 = \dots = w_N^0$. Com isto, $t = 0$ pode ser considerado como a raiz da árvore que possui N cenários.

O método de redução da árvore utilizado neste trabalho implementa um algoritmo fundamentado em uma das estratégias regressiva ou progressiva apresentadas no Capítulo 4 e no uso do princípio de redução, *erm* nos horizontes de tempo t , para cada tempo t pertencente a $\{1, 2, \dots, T\}$. Assim sendo, o algoritmo reduz sucessivamente o número de nós dos cenários $w_i = \{w_i^t\}_{t=1}^T$ (ver [8]).

¹⁴ Se w_i é uma matriz, então w_i^j é um vetor da seguinte forma: $w_i^j = \{w_i^{(1)j}, w_i^{(2)j}, \dots, w_i^{(Z)j}\}$. E matricialmente pode-se representar a árvore de cenários por uma hipermatriz $A(N, T, Z)$, onde N e T são respectivamente o número de cenários e períodos de cada elemento $z = 1, \dots, Z$. No caso de cenários que representam as afluências aos reservatórios de usinas hidrelétricas, Z representa as diferentes usinas para as quais se deseja planejar a operação conjunta. Para o SIN, valores típicos são: $1 \leq N \leq 300$; $1 \leq T \leq 12$ e $1 \leq Z \leq 320$.

Especificamente um dos algoritmos anteriormente citados é aplicado para cada instante de tempo $t = T, T - 1, \dots, 2, 1$ afim de descartar cenários cujos valores $V(P)$ e $V(Q^*)$ fiquem próximos segundo uma tolerância dada, ε_t . Como neste procedimento tem-se um conjunto de índices (cenários) a serem descartados a cada instante de tempo, chama-se este conjunto de J_t .

Antes da apresentação formal do algoritmo, o processo de redução de cenários é ilustrado pela Figura 5.5.

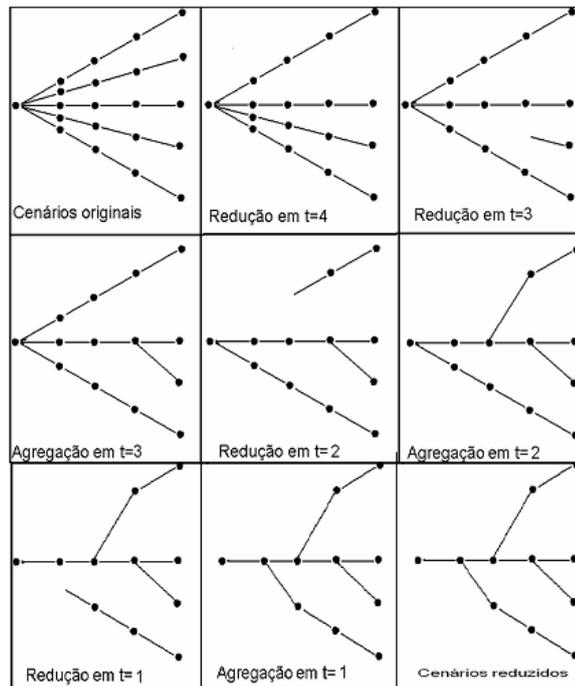


Figura 5.5: Procedimentos do Algoritmo 5.

Parâmetro(s): $(n_t, t = 1, \dots, T)$ um vetor com os números de cenários a serem preservados em cada período de tempo e/ou $(\varepsilon_t > 0, t = 1, \dots, T)$, um vetor de tolerâncias utilizado na *erm*.

Algoritmo 5 Redução da Árvore

Passo T: Aplicar a *erm* com um dos Algoritmos 2, 3 ou 4, para determinar o conjunto de índices $J_T \subset \{1, 2, \dots, N_T\} =: I_T$ tal que:

$$\sum_{i \in J_T} p_i \min_{j \in I_T \setminus J_T} \mu(w_i, w_j) \leq \varepsilon_T.$$

Aplicar a redistribuição dada pela regra (4.7).

Passo t: Para $t = T - 1, T - 2, \dots, 1$ aplicar *erm* para determinar o conjunto de

índices $J_t \subset \{1, 2, \dots, N_t\} =: I_t$ tal que $\sum_{i \in J_t} p_i \min_{j \in I_t \setminus J_t} \mu(w_i, w_j) \leq \varepsilon_t$.

Aplicar a redistribuição dada pela regra (4.7).

Agregação: Para cada $j \in J_t$, seleciona-se um índice i^* tal que

$i^* = \arg \min_{i \in I_t \setminus J_t} \mu(w_i, w_j)$ e toma-se $w_j^\tau = w_{i^*}^\tau$ para $\tau = 2, \dots, t$ (como mostra a Figura 5.5)

Passo 0: Fim de tarefa. O número de cenários, em todos períodos, foram reduzidos.

Cada passo do Algoritmo 5 reduz a árvore de cenários em diferentes períodos. No Passo T , os cenários são considerados em toda sua extensão, ou seja, do período 1 até o período T . No Passo $T-1$, os cenários preservados são considerados do período 1 até o período $T-1$, e assim sucessivamente.

Para a árvore de cenários da Figura 5.4, por exemplo, as diferentes representações matriciais da árvore considerada para cada período de tempo são:

$$A_2 = \begin{bmatrix} a & b & d \\ a & b & e \\ a & c & f \\ a & c & g \end{bmatrix}, \quad A_1 = \begin{bmatrix} a & b \\ a & c \end{bmatrix}, \quad A_0 = [a]$$

Com este exemplo, o Algoritmo 5 no Passo T trabalha com a matriz A_2 (utilizando a emr); no Passo $T-1$, trabalha com a matriz A_1 encerrando a redução de cenários com a matriz A_0 (o elemento a é determinístico).

A fase da agregação do algoritmo é necessária para não perder os nós que ficaram "órfãos" após a erm (por exemplo, como acontece no gráfico "Redução t=3" da Figura 5.5).

Suponha que o algoritmo descarte os cenários w_2 e w_3 no Passo T , deste modo, a nova matriz de cenários se torna

$$A_2 = \begin{bmatrix} a & b & d \\ a & c & g \end{bmatrix}.$$

Suponha agora que no Passo $T-1$ o cenário w_2 (linha 2) de A_1 seja descartado. Assim, a árvore reduzida toma a seguinte forma matricial $A_2 = \begin{bmatrix} a & b & d \\ a & b & g \end{bmatrix}$. Note que, se no Passo $T-1$ não fosse realizada a agregação, o elemento g seria perdido.

Uma opção para se reduzir a árvore de cenários, seria efetuar apenas o Passo T do Algoritmo 5, como apresentado no Algoritmo 6 a seguir.

Algoritmo 6 Redução da Árvore no Período T

Passo 1: Aplicar a erm com um dos Algoritmos 2, 3 ou 4, para determinar o conjunto de índices $J \subset \{1, 2, \dots, N\} =: I$ tal que:

$$\sum_{i \in J} p_i \min_{j \in I \setminus J} \mu(w_i, w_j) \leq \varepsilon.$$

Aplicar a redistribuição dada pela regra (4.7).

Assim sendo, não há a necessidade de agregação; no entanto, para os períodos $t = 1, \dots, T - 1$, não haveria descartes de cenários, salvo as reduções indiretas causadas no estágio $t - 1$ pela eliminação do cenário correspondente no estágio t . Esta é a variante aplicada aos exemplos do Capítulo 6. Pois sendo as árvores de vazões afluentes às usinas hidroelétricas geradas pelo modelo GEVAZP, que utiliza um modelo da família *ARIMA*, a fase da agregação se apresenta imprópria.

5.2.2 – Dificuldades Computacionais

O armazenamento das informações de forma eficiente é de fundamental importância para a ROC. Por exemplo, a Figura 5.4 representa uma árvore de 4 cenários que pode ser armazenada não matricialmente, mas de modo vetorial reduzido, da seguinte forma:

$$V = [b, c, d, e, f, g].$$

O elemento a é comum a todos cenários, e além disso é determinístico. Deste modo, não é necessário armazenar a em V .

Em algumas situações são aplicadas com frequência árvores que abrem a mesma quantidade de galhos em cada nó. Este número é chamado de *multiplicidade* da árvore. Exemplos comuns são árvores binárias, como multiplicidade igual a 2 ($mult = 2$, ver Figura 5.4), e árvores ternárias com multiplicidade igual 3 ($mult = 3$, ver Figura 5.6).

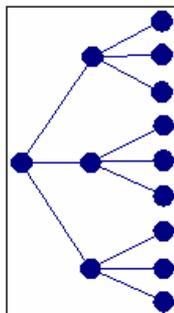


Figura 5.6: Árvore ternária.

Para este tipo de árvores, a cardinalidade do vetor V é dada por $\#V = \frac{mult^{T+1} - mult}{mult - 1}$.

O armazenamento das distâncias $\mu(w_i, w_j)$ entre os pares de cenários pode-se efetuar também em forma vetorial reduzida:

$$M_T = \begin{bmatrix} \mu(w_1, w_2) \\ \mu(w_1, w_3) \\ \mu(w_1, w_4) \\ \mu(w_2, w_3) \\ \mu(w_2, w_4) \\ \mu(w_3, w_4) \end{bmatrix}, \text{ com } \#M_T = \frac{N(N-1)}{2}.$$

porque $\mu(w_i, w_j) = \mu(w_j, w_i)$.

Suponha que se tenha uma árvore multivariada, ou seja, com $w_i^t \in \mathbb{R}^k$. Assim, $V \in \mathbb{R}^k \times \mathbb{R}^{\frac{mult^{T+1} - mult}{mult - 1}}$. Se, por exemplo, a árvore multivariada assumir $k=20$, multiplicidade igual a 30 e 5 períodos, tem-se que:

$$V \in \mathbb{R}^{20} \times \mathbb{R}^{25137930} \text{ (}\#V \text{ aproximadamente igual a } 5.1 \cdot 10^8 \text{);}$$

$$N = 5^{30} = 9.3132 \cdot 10^{20} \text{ e } M_T \in \mathbb{R}^{4.336810^{41}}.$$

Claramente, a manipulação e armazenamento deste volume de informação se torna muito complexa, se não inviável.

Ainda vale ressaltar que, para algumas aplicações o armazenamento da matriz de distâncias no vetor M_T pode impossibilitar a utilização do Algoritmo 4; visto que este depende da matriz completa, característica que é então perdida em M_T .

A seguinte seção fornece uma alternativa de redução de cenários que dispensa o armazenamento completo dos dados, já que trabalha gerando e reduzindo cenários simultaneamente.

5.2.3 – Redução Local

A Redução Local é uma forma diferenciada de construir uma árvore de cenários, com pequeno porte, que represente de forma acurada o processo estocástico. A construção se dá por meio de sucessivas gerações e reduções de cenários. Seu objetivo, assim como o Algoritmo 5, é lidar com o problema de alta dimensionalidade mencionado na seção precedente.

Esta estratégia é conveniente computacionalmente em casos em que se gera a árvore de alta dimensionalidade pretendendo reduzi-la posteriormente.

Suponha que se deseje gerar uma árvore de cenários de 12 períodos com 1000 aberturas à cada nó. Deste modo, tem-se um número total de cenários igual a 1000^{12} impossível de ser armazenado ou manipulado para efetuar a redução de cenários. A redução local permite trabalhar indiretamente com todas estas informações, usando aproximações da árvore.

A geração de uma árvore de cenários através da estratégia de redução local se fundamenta na geração de N cenários para o primeiro período de tempo e, na imediata aplicação da erm. Para cada cenário preservado, geram-se mais N cenários e aplica-se a erm em cada ramo desta árvore de dois estágios de tempo. Para cada cenário preservado no segundo período, gera-se cenários e, mais uma vez, aplica-se a erm. Este procedimento é repetido até $t = T$.

A Figura 5.7 ilustra o procedimento para a geração da árvore de cenários através da redução local para $T = 2$ estágios de tempo.

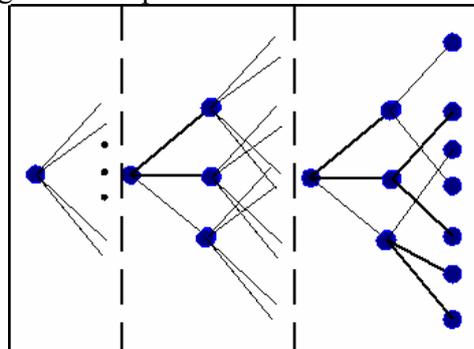


Figura5.7: Redução local.

Parâmetro(s): N_{red}^t , o número de cenários a serem preservados, satisfazendo $N_{red}^t < N^t$, e/ou $\varepsilon_t > 0$, um vetor de tolerâncias utilizado na erm para $t = 1, 2, \dots, T$.

Algoritmo 7 Redução Local.

Passo 0: Fazer $t = 1$

Passo 1: Geração de N^t cenários;

Passo t: Aplicar o Algoritmo 4 (seleção progressiva rápida) para encontrar J_t ;

Redistribuição pela regra (4.7);

Fazer $t = t + 1$; para cada nó i preservado geram-se N^t cenários, isto é, N^t filhos do nó i .

Se $t < T$, ir até o Passo t .

Se não, ir até o Passo T

Passo T: Aplicar o Algoritmo 4 para encontrar J_T ;

Redistribuir pela regra (4.7).

Desta maneira, a cada período de tempo são gerados N^t cenários que são reduzidos

a N'_{red} , de modo a construir uma árvore com cardinalidade aceitável e que represente bem as incertezas.

A escolha da estratégia progressiva como ferramenta no Algoritmo 7 permite reduzir o tempo necessário para selecionar os cenários a serem preservados. Se, por exemplo, em cada estágio de tempo geram-se 1000 cenários e deseja-se preservar apenas 5, o Algoritmo 4 é claramente mais eficiente.

A seguir uma breve apresentação das principais características dos algoritmos de seleção de cenários, e dos algoritmos de redução de árvores de cenários.

1) Redução Regressiva:

Seleciona cenários a serem descartados;

A redistribuição é realizada após o descarte de k_j cenários, com $\sum_j k_j = N - N_{red}$;

Utiliza a parte triangular inferior, ou triangular superior, da matriz de distâncias entre cenários.

2) Redução Regressia Simultânea:

Seleciona cenários a serem descartados;

A redistribuição é realizada após o descarte de $N - N_{red}$ cenários;

Utiliza a parte triangular inferior, ou triangular superior, da matriz de distâncias entre cenários.

3) Seleção Progressiva:

Seleciona cenários a serem preservados;

A redistribuição é realizada após selecionar N_{red} cenários;

Utiliza a parte triangular inferior, ou triangular superior, da matriz de distâncias entre cenários.

4) Seleção Progressiva Rápida:

Seleciona cenários a serem preservados;

A redistribuição é realizada após selecionar N_{red} cenários;

Considera a matriz de distâncias em sua forma completa.

5) Redução da árvore:

Utiliza um dos Algoritmos 2, 3 ou 4 para obter os cenários mais representativos;

Reduz a árvore em cada período, de $t = T, T - 1, \dots, 1$;

A redistribuição é realizada em cada período t ;

Realiza agregações entre os cenários;

Não recomendado para árvores de cenários obtidas a partir de modelos da família ARIMA.

6) Redução da árvore no período T :

Utiliza um dos Algoritmos 2, 3 ou 4 para obter os cenários mais representativos;

*Reduz a árvore no último período $t = T$;
A redistribuição é realizada no período $t = T$.*

7) Redução local:

*Utiliza o Algoritmo 4 para obter os cenários mais representativos;
Constrói a árvore a partir de $t = 1, 2, \dots, T$;
A redistribuição é realizada em cada período t .*

Com estas considerações, encerra-se este capítulo dedicado aos algoritmos de seleção e redução de cenários. O próximo capítulo apresenta resultados numéricos obtidos a partir de problemas reais do SIN.

6 - Resultados Numéricos

Este capítulo emprega a metodologia desenvolvida nos capítulos anteriores. São aplicados os algoritmos citados no Capítulo 5 em configurações reais do problema de planejamento da operação de curto prazo do SIN.

O planejamento da operação do SIN utiliza séries sintéticas do processo estocástico de vazões mensais afluentes aos aproveitamentos hidroelétricos obtidas pelo Modelo de Geração de Séries Sintéticas de Vazões e Energias – GEVAZP, desenvolvido pelo CEPEL, [16,17,18]. O modelo GEVAZP gera cenários de vazões mensais afluentes com um modelo auto-regressivo periódico, usualmente denotado PAR(p). Para maiores informações, ver [14,25].

De posse das informações geradas pelo modelo GEVAZP, a etapa seguinte consiste em determinar, dentre outras, as metas ótimas de geração de cada usina de um sistema hidrotérmico, sujeito a afluições incertas. O problema de otimização correspondente minimiza o valor esperado do custo de operação e déficit de energia ao longo do período de planejamento. Idealmente, para o Modelo de Determinação da Coordenação da Operação à Médio Prazo – DECOMP [19,27], deve-se ter uma representação acurada das incertezas nas vazões afluentes aos aproveitamentos hidroelétricos. Atualmente o programa DECOMP é utilizado com o horizonte de tempo de apenas 2 meses, sendo o primeiro mês discretizado em etapas semanais que consideram vazões conhecidas. Já no segundo mês, as vazões afluentes são consideradas incertas, e atualmente representadas por árvores com mais de 100 cenários. Para maior clareza e eficiência na decisão que se deve tomar no presente, deseja-se que o horizonte de tempo seja estendido. Para tanto, deve-se ter um número razoável de cenários. O objetivo é reduzir a dimensionalidade da árvore de cenários gerada pelo GEVAZP, de modo que a árvore reduzida represente adequadamente as incertezas, e que seja viável computacionalmente para o modelo DECOMP.

O procedimento atual aplicado às séries geradas pelo modelo GEVAZP seleciona um subgrupo reduzido de cenários agrupando eventos semelhantes em uma mesma classe, ou conglomerado. Para cada conglomerado, escolhe-se um elemento representante. A probabilidade de ocorrência deste representante é definida de maneira proporcional ao número de cenários que formam o conglomerado, [13].

Uma variante do processo de agregação em conglomerados, implementada no modelo GEVAZP e chamada de K-Médias (ou K-Means) [18], consiste em, para cada estágio do horizonte de estudo e para cada "ramo" da árvore de cenários, gerar (em geral 1000) novos cenários hidrológicos equiprováveis. Estes novos cenários são separados em k conglomerados, que fornecem k cenários representantes. Uma comparação entre a ROC e a agregação em conglomerados é realizada em [21]. Outro procedimento de diminuição de esforço computacional aplicado às séries de vazões afluentes aos reservatórios pode ser encontrado em [4].

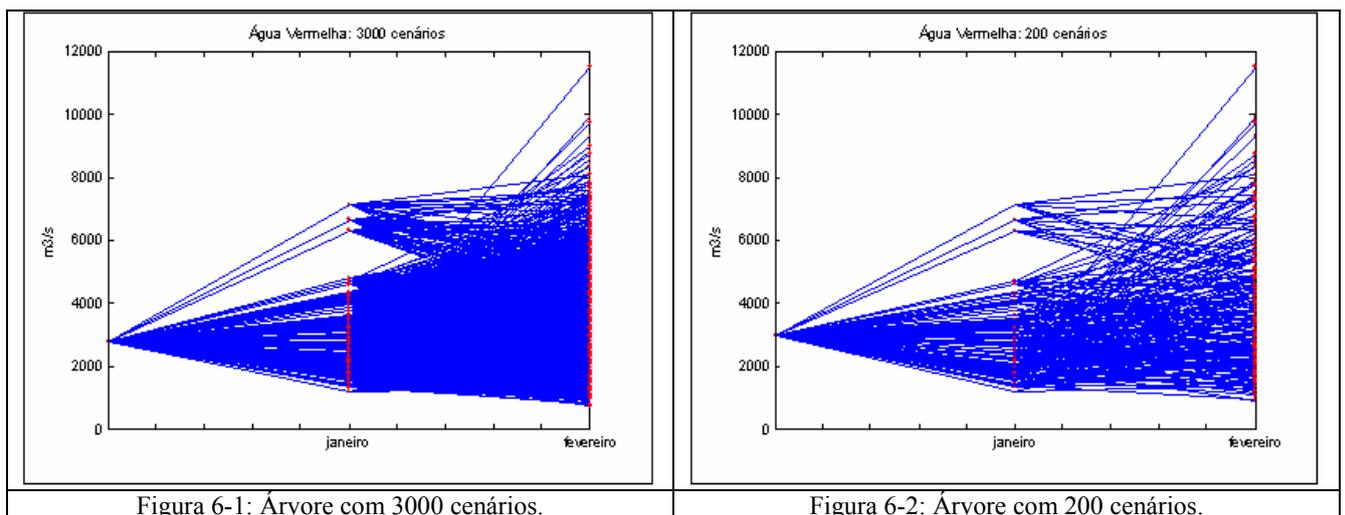
Este capítulo está dividido em 2 seções. Primeiro, se faz uma análise estatística (analisando árvores de cenários univariados e multivariados). A seguir se faz uma análise do desempenho dos diferentes algoritmos.

6.1 – Análises Estatísticas

Esta seção se dedica à comparações estatísticas entre as árvores original e reduzida. Tais análises dependem da estrutura univariada ou multivariada. Considera-se uma árvore de cenários multivariados quando se tem uma árvore de cenários (gerada multivariadamente) para cada usina pertencente a configuração estudada.

6.1.1 – Árvore de Cenários Univariados

Considere a árvore univariada de séries sintéticas de vazões de afluentes (m^3/s) que chegam a uma usina dada, apresentada na Figura 6-1. A árvore contém 3000 (150×20) cenários hidrológicos da usina hidroelétrica de Água Vermelha (Bacia do Rio Paraná), dispostos em 2 períodos temporais ($T = 2$). A Figura 6-2 apresenta a árvore de 200 cenários obtida após a redução.



No caso univariado, a distância entre os cenários pode ser dada, por exemplo, como em (3.16) com $r = 2$.

A Tabela 6-1 apresenta as médias, desvios padrões, valores (também chamados de "nós" da árvore) máximos e mínimos de cada período temporal, tanto da árvore gerada (original) como da árvore reduzida; confirmando assim, a grande acurácia do processo de ROC em árvores de cenários univariados. Neste capítulo, ao invés de utilizar o termo "árvore original", adota-se o termo "árvore gerada", porque a árvore original utilizada foi, de fato, gerada pelo modelo GEVAZP.

Tabela 6-1: Análise univariada.

Água vermelha	1º período		2º período	
	Árvore Ger.	Árvore Red.	Árvore Ger.	Árvore Red.
Médias	2749.35	2748.47	3297.75	3296.26
Desvios padrões	1019.24	1015.75	1323.50	1317.72
Valor mínimo	1161.29	1161.29	726.42	794.27
Valor Máximo	7122.46	7122.46	11493.72	11493.72

Dentre as estatísticas apresentadas na Tabela 6-1, a maior degradação causada pela redução de cenários ocorreu no desvio padrão do segundo período. Em termos percentuais, esta degradação equivale a 0.44% do valor do desvio padrão do segundo período. Uma perda significativamente pequena.

Para o problema de operação de curto prazo do SIN, a configuração utilizada consiste claramente em mais de uma usina. Por este motivo, utilizam-se árvores de cenários multivariados. A seguir os estudos do processo de ROC para este tipo de árvore.

6.1.2 – Árvore de Cenários Multivariados

Para árvores de cenários multivariados, cada cenário w_i possui uma representação matricial $w_i \in \mathbf{R}^{T \times U}$, sendo T o número de períodos e U o número de usinas utilizadas na configuração. Assim sendo, a norma utilizada na função que mede a distância entre os cenários precisa ser uma norma matricial; podendo ser, por exemplo, a função definida em (3.16) com a norma dada por

$$\|w - \tilde{w}\| = \max_{l=1,2,\dots,U} \sum_{s=1}^T |w(s,l) - \tilde{w}(s,l)|,$$

ou,

$$\|w - \tilde{w}\| = \max_{l=1,2,\dots,U} \|w(:,l) - \tilde{w}(:,l)\|_2 \quad (6.1)$$

Como no caso univariado, as médias e desvios padrões são analisados nesta seção; além disso, para árvores de cenários multivariados a análise estatística é estendida à correlação temporal e à correlação espacial entre as usinas hidroelétricas. A correlação temporal de uma usina é entendida como a correlação entre os períodos de tempo. Para a operação do SIN, a correlação entre as usinas é um fator importante no processo estocástico de geração de vazões afluentes aos reservatórios. Como as afluições dependem de condições climáticas, espera-se que usinas em uma mesma bacia hidrológica sejam positivamente correlacionadas. Este tipo de correlação é conhecido na literatura como correlação espacial.

O objetivo desta subseção é mostrar que o processo de redução de cenários preserva eficientemente as características estatísticas dos dados iniciais, incluindo as correlações espaciais e temporais.

As análises estatísticas apresentadas consideram árvores de cenários relativas a uma configuração real do SIN.

6.1.3 – Caso Estudado

Para o SIN, com um horizonte temporal de dois períodos ($T = 2$), serão apresentados a seguir estudos do caso em que a dimensionalidade da árvore de cenários gerada pelo modelo GEVAZP é reduzida em 90%. A qualidade da árvore reduzida ótima obtida é avaliada mediante diferentes testes estatísticos. Em particular, tais testes indicam uma aderência satisfatória entre o conjunto de cenários preservados na árvore reduzida, e o conjunto de cenários originais que compõem a árvore gerada.

A configuração utilizada é a seguinte:

- Histórico: 1931-1998
- Tendência hidrológica: 1959
- Mês inicial: janeiro
- Períodos: 2
- Número de cenários gerados para o primeiro período: 100
- Número de cenários gerados para o segundo período: 10000
- Número de cenários reduzidos para o primeiro período: 99
- Número de cenários reduzidos para o segundo período: 1000
- Número de usinas principais na configuração: 39
- Usinas analisadas: Tucuruí (bacia do Tocantins), Três Marias (bacia de São Francisco), Salto Caxias (bacia do Iguazu), São Simão (bacia do Rio Paranaíba), Água Vermelha (bacia do Rio Paraná), Itá (bacia do Rio Uruguai), Curuá-Una (Rio Curuá-Una), Porto Primavera (bacia do Rio Paraná) e Ilha dos Pombos (bacia do Rio Paraíba do Sul).

Para os resultados numéricos que se seguem, foi utilizado o Algoritmo 6 para a redução da árvore; sendo o Algoritmo 4 para a seleção de cenários.

A seguir as análises estatísticas entre as árvores gerada (árvore original) e reduzida.

6.1.4 – Regressão Linear: Médias e Desvios Padrões

A fim de analisar os dados reduzidos, são apresentadas na Figura 6.1 as regressões lineares entre as estatísticas das árvores gerada e reduzida de todas as 39 usinas da configuração considerada. São apresentados também o coeficiente angular da reta de regressão linear ($y = \theta x$), bem como o coeficiente de ajuste *R-square* (R^2) que mede o ajuste da reta de regressão linear às estatísticas dos valores gerados, para maiores informações consultar a referência [9].

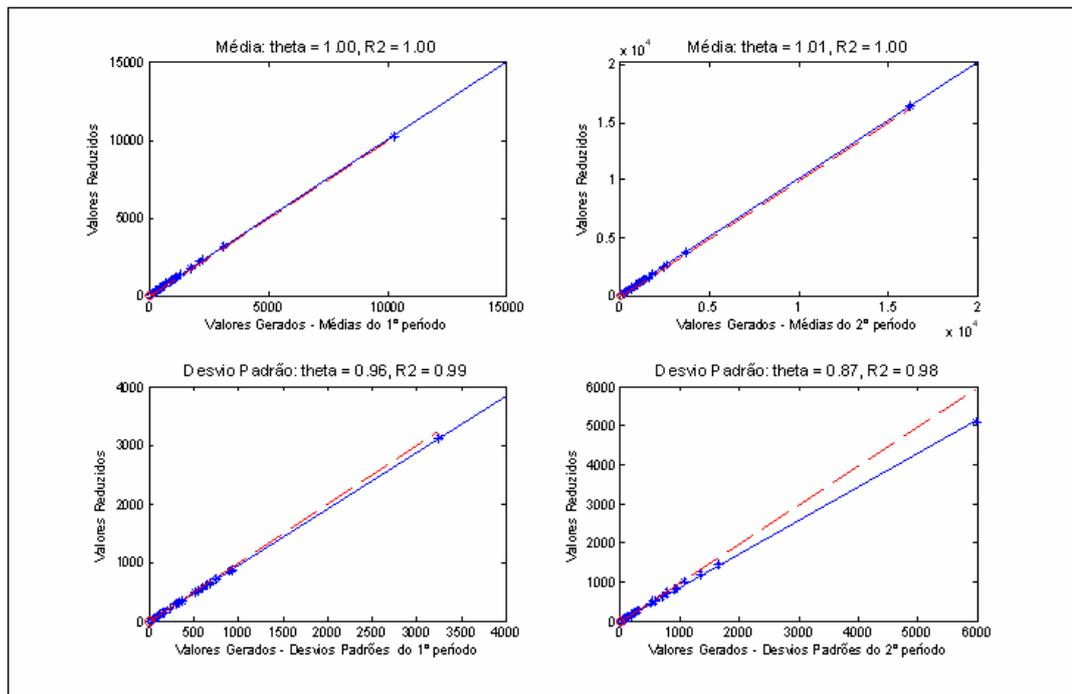


Figura 6.1: Regressão linear entre as médias e desvios padrões - 1º e 2º períodos.

Como todo desenvolvimento da técnica de ROC é fundamentado na minimização da diferença entre os valores esperados (expressão (4.3)), os excelentes resultados em relação à estatística média confirmado pela Figura 6.1 (parte superior) eram previsíveis. A parte inferior da Figura 6.2 mostra que também os desvios padrões foram eficientemente preservados em ambos os períodos. Estes resultados são confirmados pelo coeficiente angular da reta de regressão linear, e pelo coeficiente de determinação R^2 .

6.1.5 – Regressão Linear: Correlação Temporal

Um elemento importante para a análise é a correlação entre as aflúências dos períodos 2 e 1 de cada usina hidroelétrica. A Figura 6.2 compara esta estatística, quando obtida com os cenários gerados e pela ROC para cada uma das usinas da configuração.

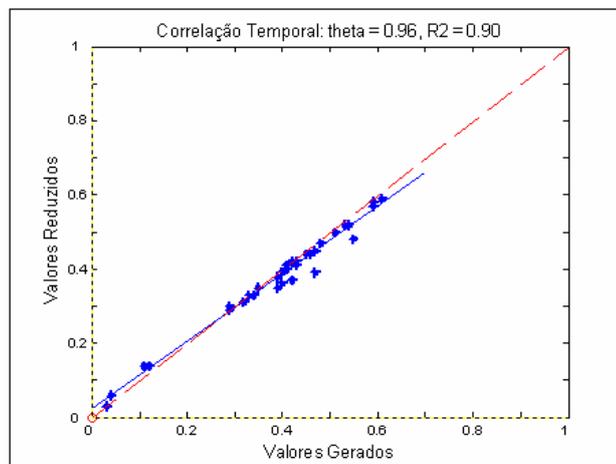


Figura 6.2: Correlação temporal.

Como se pode verificar, a Figura 6.2 apresenta bons resultados de aderência entre as correlações temporais (1º e 2º períodos) das usinas.

Como proposto, serão apresentados a seguir as médias e desvios padrões de 9 usinas da configuração utilizada.

6.1.6 – Valores Numéricos das Usinas Estudadas

São apresentados nesta subseção tabelas contendo as médias, desvios padrões, valores máximos e mínimos de cada período, tanto da árvore gerada como da árvore reduzida. Também são apresentados valores teóricos¹⁵ das médias e desvios padrões.

Assim como as médias e desvios padrões, a correlação espacial entre as usinas são apresentadas graficamente na subseção 6.1.3, onde se encontram também os testes de aderência das distribuições de probabilidades para cada uma das 9 usinas analisadas.

A tabela seguinte apresenta os resultados obtidos pela redução de cenários para as usinas de Tucuruí, Três Marias e Salto Caxias.

Tabela 6-2: Análise multivariada - Tucuruí, Três Marias e Salto Caxias.

	1º período			2º período		
	Gerado	Reduzido	Teórico	Gerado	Reduzido	Teórico
Usina: Tucuruí						
Médias	10266.35	10247.29	10875.26	16266.81	16462.60	16852.83
Desvios padrões	3247.53	3121.43	3676.38	5972.40	5104.78	6177.66
Valor mínimo	3906.97	3906.97	/	1789.26	4502.70	/
Valor Máximo	22226.96	22226.96	/	53111.17	53111.17	/
Usina: Três Marias						
Médias	1317.41	1345.51	1270.44	1305.30	1347.14	1295.06
Desvios padrões	646.65	635.95	674.32	791.75	744.68	804.82

¹⁵ A média e desvio padrão teórico correspondem aos valores esperados do modelo auto-regressivo.

	1º período			2º período		
Valor mínimo	292.03	292.03	/	69.45	235.07	/
Valor Máximo	3455.19	3455.19	/	16016.23	16016.23	/
Usina: Salto Caxias						
Médias	906.83	931.53	880.82	1125.49	1144.53	1116.40
Desvios padrões	684.10	650.74	762.19	710.73	630.00	727.87
Valor mínimo	90.10	90.10	/	95.41	109.05	/
Valor Máximo	3855.01	3855.01	/	7112.89	7112.89	/

Como mostra a Tabela 6-2, no 2º período, a distância entre a média e o valor mínimo é muito menor que a distância entre a média e o valor máximo. Por esta razão, e pelo fato de que os cenários são equiprováveis, pode-se concluir que valores inferiores à média são em maior número do que os valores superiores a ela. Como neste caso a redução de cenários foi rigorosa (redução de 90% do total de cenários) é natural que o descarte de cenários seja mais intenso nos valores inferiores à média. Isso explica o fato pelo qual o valor máximo foi preservado, e o valor mínimo foi descartado (não necessariamente em todas as usinas). O principal fator que influencia na escolha de cenários é a função que mede a distância entre eles. Neste estudo foi utilizada a função μ como em (6.1).

A seguir a Tabela 6-3 apresenta resultados para mais 4 usinas.

Tabela 6-3: Análise multivariada - São Simão, Água Vermelha, Itá e Curua-Una.

	1º período			2º período		
Usina: São Simão	Gerado	Reduzido	Teórico	Gerado	Reduzido	Teórico
Médias	2251.73	2283.82	2289.59	2620.59	2692.48	2709.84
Desvios padrões	759.08	735.86	824.20	1068.90	1015.87	1146.83
Valor mínimo	1135.90	1135.90	/	348.12	768.99	/
Valor Máximo	5191.12	5191.12	/	8921.65	8921.65	/
Usina: Á. Vermelha						
Médias	728.11	740.02	745.57	1065.34	1081.98	1081.92
Desvios padrões	310.98	303.60	346.17	515.05	466.17	537.91
Valor mínimo	174.20	174.20	/	141.78	247.62	/
Valor Máximo	1741.07	1741.07	/	4912.40	4912.40	/
Usina: Itá						
Médias	473.55	491.42	470.80	693.73	698.93	692.93
Desvios padrões	294.79	284.50	373.96	582.14	530.07	608.46
Valor mínimo	59.66	59.66	/	8.62	36.31	/
Valor Máximo	1692.97	1692.97	/	11888.67	11888.67	/
Usina: Curua-Una						
Médias	106.68	104.94	107.85	162.35	160.54	162.08
Desvios padrões	26.82	25.67	26.30	52.40	43.32	51.83
Valor mínimo	57.74	57.74	/	35.19	51.44	/
Valor Máximo	221.04	221.04	/	434.84	378.85	/

Assim como na Tabela 6-2, as médias e desvios padrões dos cenários reduzidos estão, respectivamente, próximos às médias e desvios padrões dos cenários gerados. Estes últimos, por sua vez, são aderentes aos valores teóricos.

Tabela 6-4: Análise multivariada - Porto primavera e Ilha dos Pombos.

	1º período			2º período		
	Gerado	Reduzido	Teórico	Gerado	Reduzido	Teórico
Usina: P. Primavera						
Médias	2122.88	2182.55	2205.02	2475.73	2508.13	2476.77
Desvios padrões	582.69	548.84	661.85	925.69	804.99	953.47
Valor mínimo	912.73	912.73	/	567.17	829.60	/
Valor Máximo	3816.60	3816.60	/	8814.04	8814.04	/
Usina: I. Pombos						
Médias	316.78	325.02	297.97	331.04	334.13	326.36
Desvios padrões	162.04	158.04	157.06	176.94	158.85	180.46
Valor mínimo	70.61	70.61	/	23.57	59.55	/
Valor Máximo	987.30	987.30	/	2130.97	2130.97	/

As Tabelas 6-2, 6-3, e 6-4 apresentam as médias e desvios padrões teóricos. As estatísticas dos cenários gerados (ou cenários originais) e as estatísticas dos cenários reduzidos se comportam de maneira satisfatória em relação às estatísticas teóricas. Uma análise gráfica destes resultados é apresentada a seguir.

6.1.7 – Resultados - Figuras

Uma primeira avaliação da qualidade dos cenários preservados consiste em comparar as estatísticas da árvore original (a árvore gerada pelo modelo GEVAZP) com os valores teóricos, segundo descritos em [5]. A árvore reduzida pela qual são apresentados os valores estatísticos a seguir possui 1000 cenários, que equivalem a 10% do total de cenários gerados.

A média teórica e o desvio padrão teórico correspondem, respectivamente, ao valor esperado e ao desvio padrão do modelo auto-regressivo periódico, condicionado ao passado. A Figura 1 a seguir contém as médias e desvios padrões dos dados gerados e os valores teóricos.

Médias e Desvios Padrões da Usina de Tucuruí

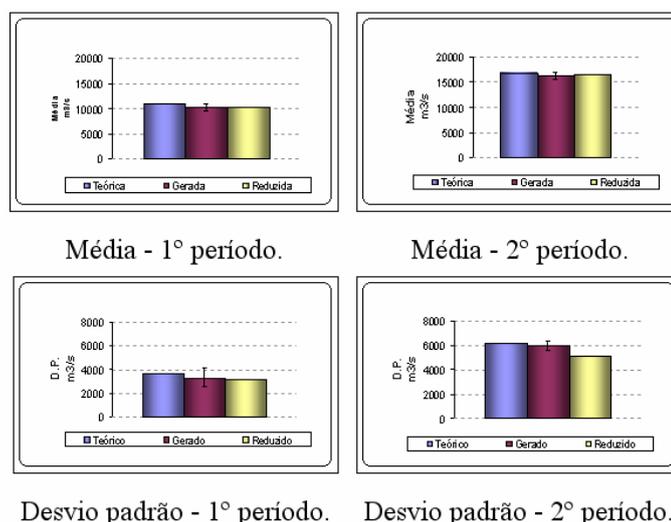


Figura 1: Médias e Desvios Padrões - Tucuruí

A Figura 1 mostra que as estatísticas dos valores gerados estão contidas nos intervalos de aceitabilidade representados pelas barras de erros. As estatísticas dos cenários gerados (ou cenários originais) e as estatísticas dos cenários reduzidos pela técnica ROC se comportam de maneira satisfatória em relação aos valores teóricos.

Vale ressaltar que para comparar as estatísticas reduzidas com as estatísticas teóricas os intervalos de aceitabilidade seriam mais amplos (a amplitude do intervalo de aceitabilidade é inversamente proporcional à raiz quadrada do tamanho da amostra).

Correlação Espacial da Usina de Tucuruí

A Figura 2 apresenta as correlações espaciais entre as vazões da usina hidroelétrica de Tucuruí com as demais usinas hidroelétricas da configuração. A correlação espacial determina as dependências hidrológicas entre as usinas. Como as aflúncias aos

reservatórios dependem das condições climáticas, usinas de uma mesma bacia hidrológica são positivamente correlacionadas.

O eixo das abscissas correspondem às correlações espaciais obtidas a partir dos cenários gerados. O eixo das ordenadas correspondem às correlações espaciais obtidas a partir dos cenários reduzidos.

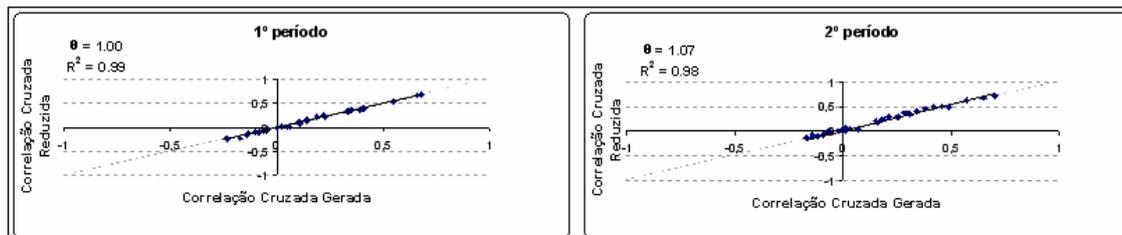


Figura 2: Regressão linear entre as correlações espaciais – 1º e 2º períodos.

Para este estudo, os cenários reduzidos representam apenas 10% do total de cenários gerados. No entanto, como mostra a Figura 2, este número bem reduzido de cenários ainda reproduz eficientemente as correlações espaciais entre as usinas hidrológicas da configuração.

Testes de Aderência - Tucuruí

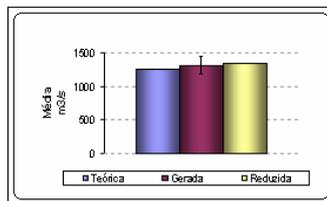
A tabela seguinte fornece os valores críticos referentes aos testes de aderência de Kolmogorov-Smirnov e Cramér-von Mises [26] para as distribuições de cada usina. Estes testes medem a aderência entre duas distribuições de probabilidades. Se o valor obtido for inferior ao valor crítico, considera-se que as distribuições de probabilidades, referentes à árvore original e à árvore reduzida, são aderentes.

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.626	0.413
Cramér – Von Mises valor crítico: 0.46 (5%)	0.057	0.002

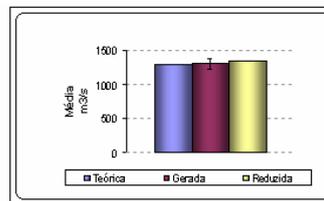
Os valores da tabela acima confirmam mais uma vez a eficiência da técnica de ROC. Os valores obtidos são significativamente inferiores aos limites de aceitação. Este resultado garante a aderência entre as distribuições de probabilidades P e Q^* de (4.7).

Para o estudo das próximas 8 usinas, as conclusões são análogas às da usina de Tucuruí.

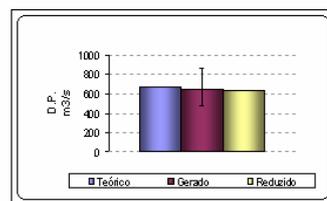
Médias e Desvios Padrões da Usina de Três Marias



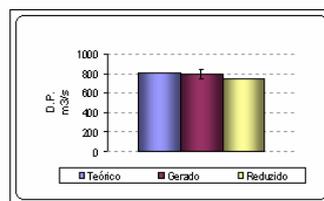
Média - 1º período.



Média - 2º período.

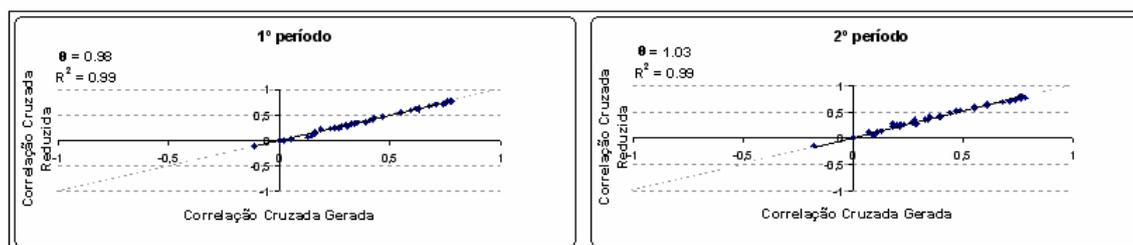


Desvio padrão - 1º período.



Desvio padrão - 2º período.

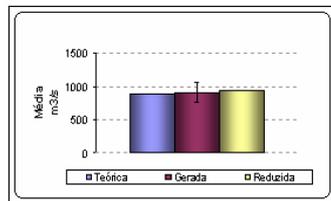
Correlação Espacial da Usina de Três Marias



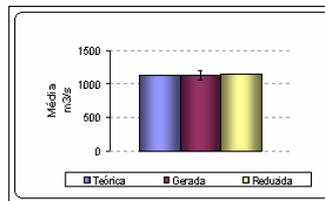
Testes de Aderência - Três Marias

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.372	0.885
Cramér – Von Mises valor crítico: 0.46 (5%)	0.018	0.056

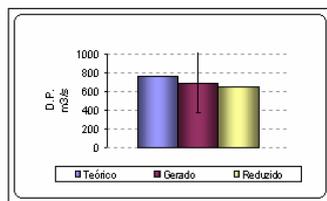
Médias e Desvios Padrões da Usina Salto Caxias



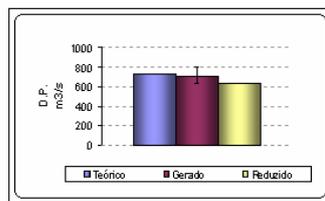
Média - 1º período.



Média - 2º período.

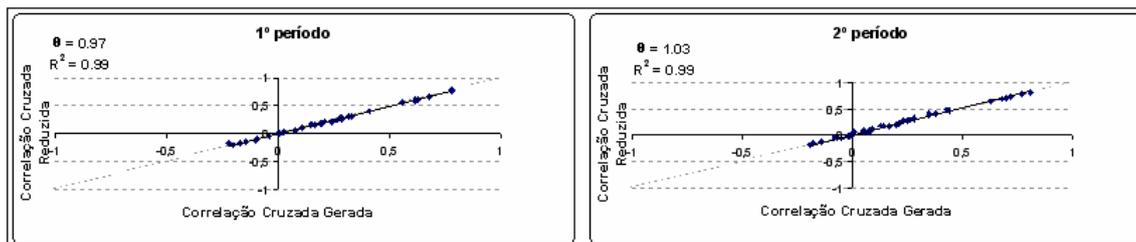


Desvio padrão - 1º período.



Desvio padrão - 2º período.

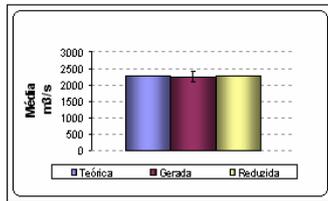
Correlação Espacial da Usina Salto Caxias



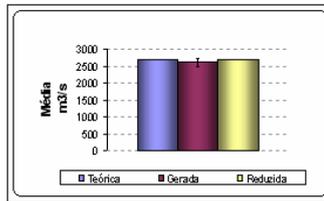
Testes de Aderência - Salto Caxias

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.774	0.792
Cramér – Von Mises valor crítico: 0.46 (5%)	0.128	0.022

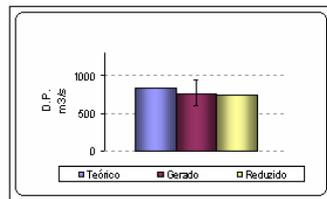
Médias e Desvios Padrões da Usina São Simão



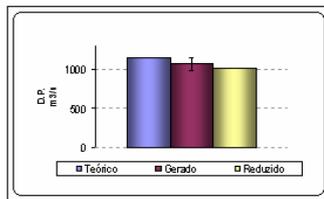
Média - 1º período.



Média - 2º período.

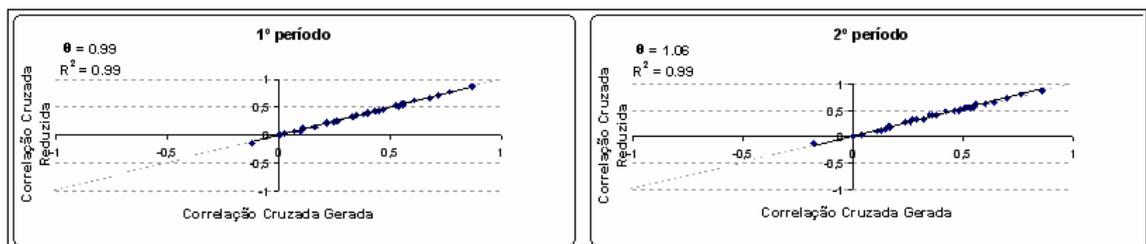


Desvio padrão - 1º período.



Desvio padrão - 2º período.

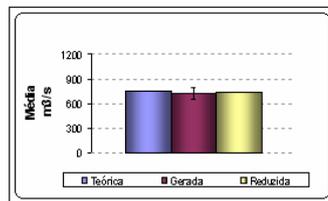
Correlação Espacial da Usina São Simão



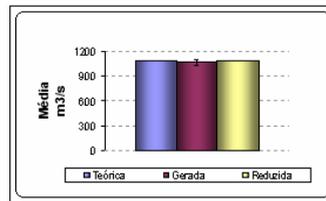
Testes de Aderência - São Simão

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.589	0.467
Cramér – Von Mises valor crítico: 0.46 (5%)	0.054	0.01

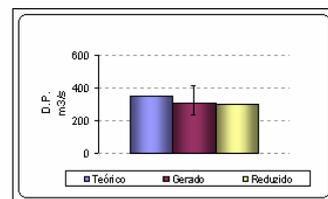
Médias e Desvios Padrões da Usina Água Vermelha



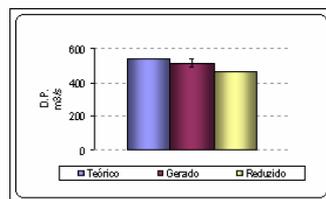
Média - 1º período.



Média - 2º período.

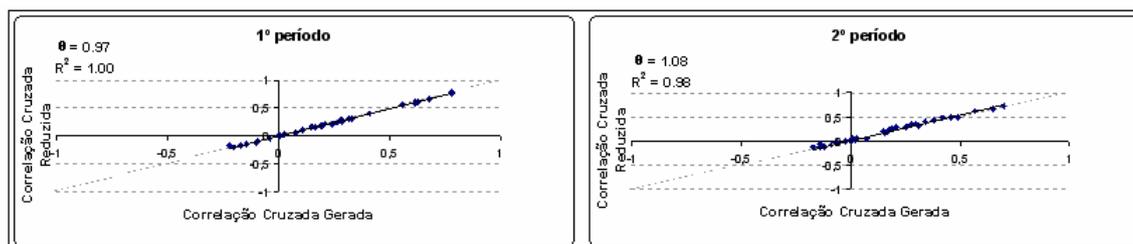


Desvio padrão - 1º período.



Desvio padrão - 2º período.

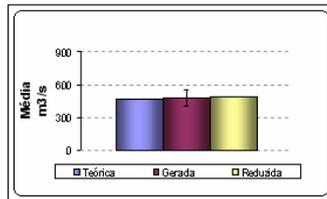
Correlação Espacial da Usina Água Vermelha



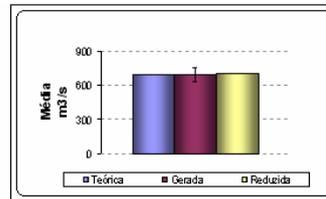
Testes de Aderência - Água Vermelha

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.358	0.477
Cramér – Von Mises valor crítico: 0.46 (5%)	0.011	0.008

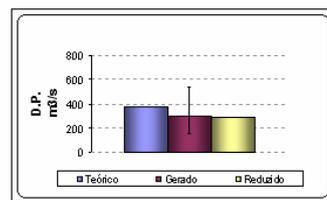
Médias e Desvios Padrões da Usina Itá



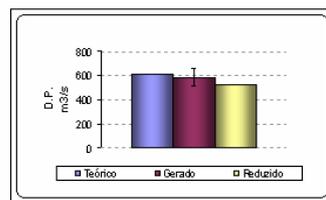
Média - 1º período.



Média - 2º período.

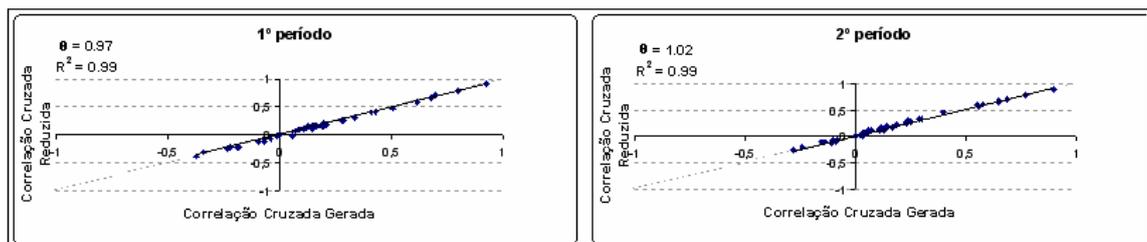


Desvio padrão - 1º período.



Desvio padrão - 2º período.

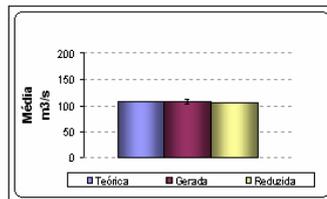
Correlação Espacial da Usina Itá



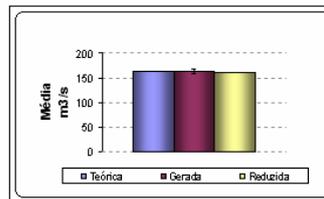
Testes de Aderência - Itá

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.593	0.649
Cramér – Von Mises valor crítico: 0.46 (5%)	0.085	0.008

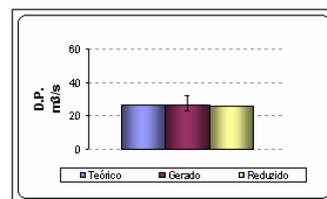
Médias e Desvios Padrões da Usina Curua-Una



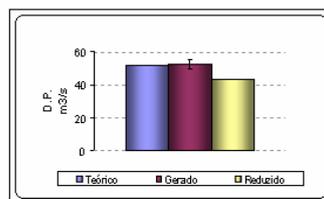
Média - 1º período.



Média - 2º período.

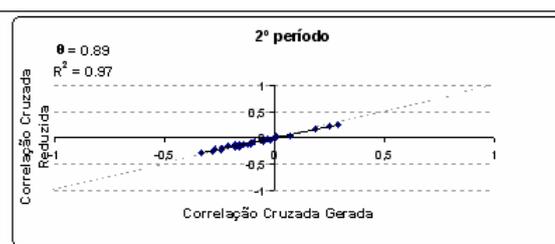
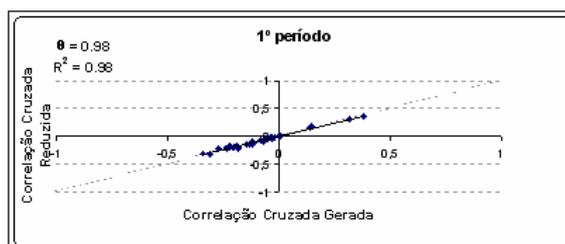


Desvio padrão - 1º período.



Desvio padrão - 2º período.

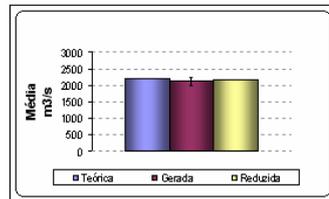
Correlação Espacial da Usina Curua-Una



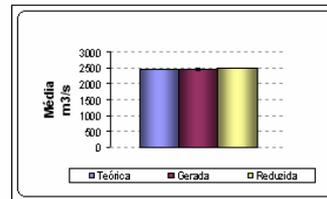
Testes de Aderência - Curua-Una

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.44	0.481
Cramér – Von Mises valor crítico: 0.46 (5%)	0.076	0.061

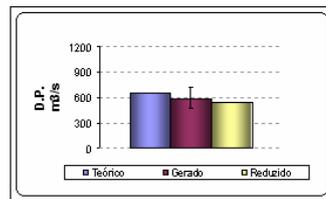
Médias e Desvios Padrões da Usina Porto Primavera



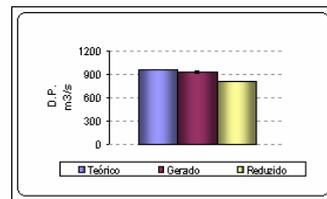
Média - 1º período.



Média - 2º período.

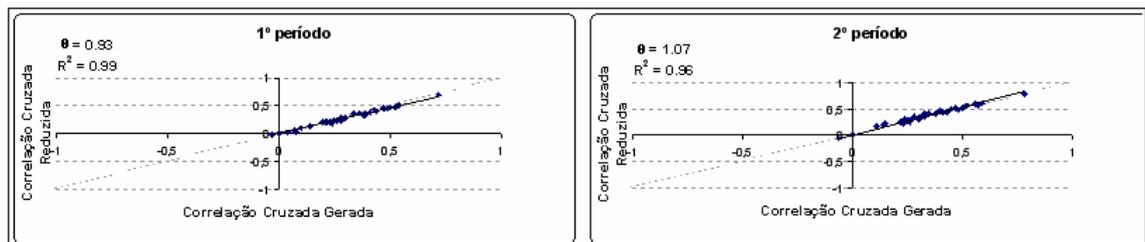


Desvio padrão - 1º período.



Desvio padrão - 2º período.

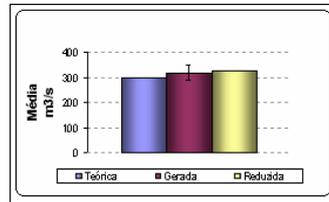
Correlação Espacial da Usina Porto Primavera



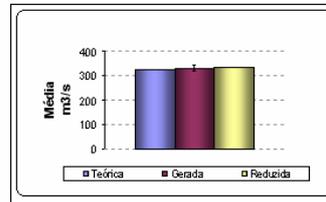
Testes de Aderência - Porto Primavera

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.613	0.585
Cramér – Von Mises valor crítico: 0.46 (5%)	0.124	0.022

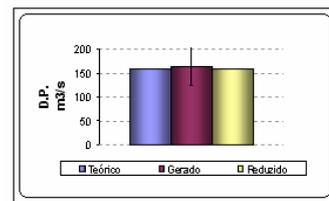
Médias e Desvios Padrões da Usina Ilha dos Pombos



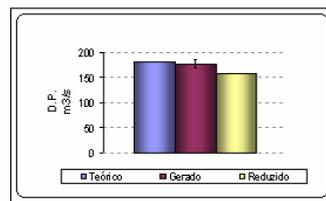
Média - 1º período.



Média - 2º período.

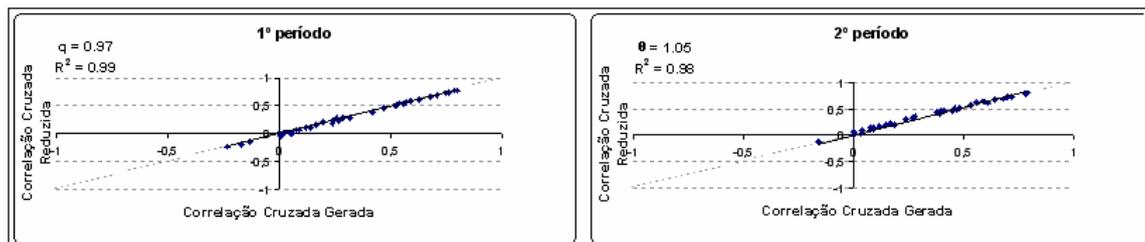


Desvio padrão - 1º período.



Desvio padrão - 2º período.

Correlação Espacial da Usina Ilha dos Pombos



Testes de Aderência - Ilha dos Pombos

Teste de aderência	1º período	2º período
Kolmogorov – Smirnov valor crítico: 1.36 (5%)	0.504	0.575
Cramér – Von Mises valor crítico: 0.46 (5%)	0.062	0.011

Melhores resultados para o desvio padrão foram obtidos quando se utiliza a norma (6.1) na função $\mu(w, \tilde{w}) = \|w - \tilde{w}\| \max\{1, h(\|w\|), h(\|\tilde{w}\|)\}$, onde a função h é definida como $h(\|w\|) = \|w\|^{2(1+e^{pc(l^*)})}$, sendo $pc(l^*)$ o "percentual de contribuição"¹⁶ da usina l^* da configuração estudada. A usina l^* é tal que

¹⁶ Sejam $l = 1, \dots, U$ as usinas da configuração. Seja $\varphi(l)$ a média dos cenários gerados para cada usina l . Define-se, neste trabalho, o percentual de contribuição da usina l a seguinte razão: $pc(l) := \frac{\varphi(l)}{\sum_{i=1}^U \varphi(i)}$.

$$\|w - \tilde{w}\| = \max_{l=1,2,\dots,U} \|w(:,l) - \tilde{w}(:,l)\|_2 = \|w(:,l^*) - \tilde{w}(:,l^*)\|_2.$$

É importante que os cenários estejam padronizados, pois do contrário a usina que possuir maior potencial hidrológico causará um viés no cálculo das distâncias entre cenários (l^* permanece fixo na maioria dos cálculos de $\mu(w, \tilde{w})$).

A seguir uma análise dos algoritmos de seleção de cenários apresentados no Capítulo 5.

6.2 – Análise do Desempenho dos Algoritmos

Esta subseção destina-se a tratar de forma mais concreta os três diferentes algoritmos de seleção, baseados nas duas modalidades apresentadas (Regressiva e Progressiva). Para deixar clara a eficiência de tais algoritmos em diversas condições de redução, será apresentada a seguir uma tabela com as medidas (em percentual) da proximidade entre os dados gerados e reduzidos e, contendo também, os tempos de processamento para ROC consistindo de seiscentos cenários gerados.

Para calcular (em termos percentuais) a distância entre a medida de probabilidade inicial (i.e., dada pela árvore gerada com cenários equiprováveis) e a medida de probabilidade reduzida (árvore reduzida e redistribuída de forma ótima) é utilizada a *medida de precisão relativa* Δ definida a seguir, de modo a proporcionar um valor percentual da ROC.

$$\Delta(N_{red}) := \frac{c(J)}{c(\{i^*\})},$$

onde $i^* \in \{1, \dots, N\}$ denota o primeiro índice de cenário obtido pela seleção progressiva, ver Exemplo 3. Deste modo, $c(\{i^*\})$ corresponde à possível melhor distância de P a um único cenário w_{i^*} , que neste caso, possui 100% de probabilidade de ocorrência.

A Tabela 6-5 a seguir apresenta o tempo de processamento e a medida de precisão relativa dos Algoritmos 2, 3 e 4.

Vale ressaltar que o tempo de processamento apresentados na seguinte tabela não incluem o tempo de geração de cenários pelo modelo GEVAZP.

Tabela 6-5: Análise do desempenho dos algoritmos.

Nº de cenários:		R. R. S. (alg. 2)		S. P. (alg. 3)		S. P. R. (alg. 4)	
N	N_{red}	tempo	$\Delta(N_{red})$	tempo	$\Delta(N_{red})$	tempo	$\Delta(N_{red})$
600	100	42' 21"	25,8	10' 12"	25,1	0' 23"	25,1
600	200	37' 30"	17,9	24' 40"	17,4	0' 24"	17,4
600	300	30' 18"	12,1	38' 53"	11,7	0' 24"	11,7
600	400	15' 40"	7,2	47' 10"	7,1	0' 26"	7,1
600	500	5' 4"	3,2	60' 35"	3,2	0' 27"	3,2

A Tabela 6-5 mostra que o descarte de 50% dos cenários implica em uma perda de aproximadamente 12% da precisão dos dados. Além disso, no exemplo ilustrado nesta tabela é possível encontrar uma sub-árvore com seis cenários (1%) que acarreta uma perda de 49,4% da precisão relativa. Vale ressaltar que melhores resultados para a medida de precisão relativa podem ser obtidos no caso em que a árvore é univariada. Este resultado pode ser verificado em [10].

O Algoritmo 2 (Redução Regressiva Simultânea) obteve medidas de precisão relativa um pouco melhores que os algoritmos da modalidade progressiva. Este fato se deve ao modo de como cada algoritmo procede no caso de empate na escolha de determinados cenários. O empate de cenários é considerado como sendo, na iteração k -ésima do algoritmo, $c(J^{k-1} \cup i) = c(J^{k-1} \cup j)$ para diferentes índices de cenários i, j . Neste sentido, distintos algoritmos podem selecionar diferentes índices e, por esta razão, obter conjuntos J um pouco diferentes. Uma maneira de se escolher os índices empatados, é preservar aquele que está mais distante da média. Pois deste modo, existe a possibilidade de se contribuir para a preservação do desvio padrão.

Como mencionado no Capítulo 5, o Algoritmo 2 é bastante eficaz quando se quer realizar uma redução fraca. Por outro lado, o Algoritmo 3 possui melhor rendimento no caso da redução forte. No entanto, ambos algoritmos são inferiores ao Algoritmo 4 (Seleção Progressiva Rápida) no que se diz respeito ao tempo de processamento.

A Tabela 6-5 mostra que o Algoritmo 4 é superior ao Algoritmo 3. Pois como foi provado no Teorema 6, estes são equivalentes no que diz respeito aos índices de cenários preservados, porém, devido a complexidade computacional, o Algoritmo 4 é significativamente mais eficiente. A complexidade deste algoritmo decresce a medida que o número de cenários preservados (N_{red}) decresce; para mais detalhes sobre a complexidade de cada algoritmo, consulte as proposições 2.3 e 2.6 em [8].

6.2.1 – Análise do Algoritmo de Redução Local

Como apresentado no Capítulo 5, a Redução Local é uma forma de construir uma árvore de cenários por meio de sucessivas gerações e reduções de cenários.

A Figura 6.3 apresenta os resultados de regressão linear entre as médias e desvios padrões obtidos a partir de uma árvore com 2 períodos e 100 cenários (10 aberturas para o 1º período e 10 aberturas para o 2º período) construída utilizando o Algoritmo 7. Para obter o subconjunto de 100 cenários, foram selecionados 10 a cada 1000 cenários gerados.

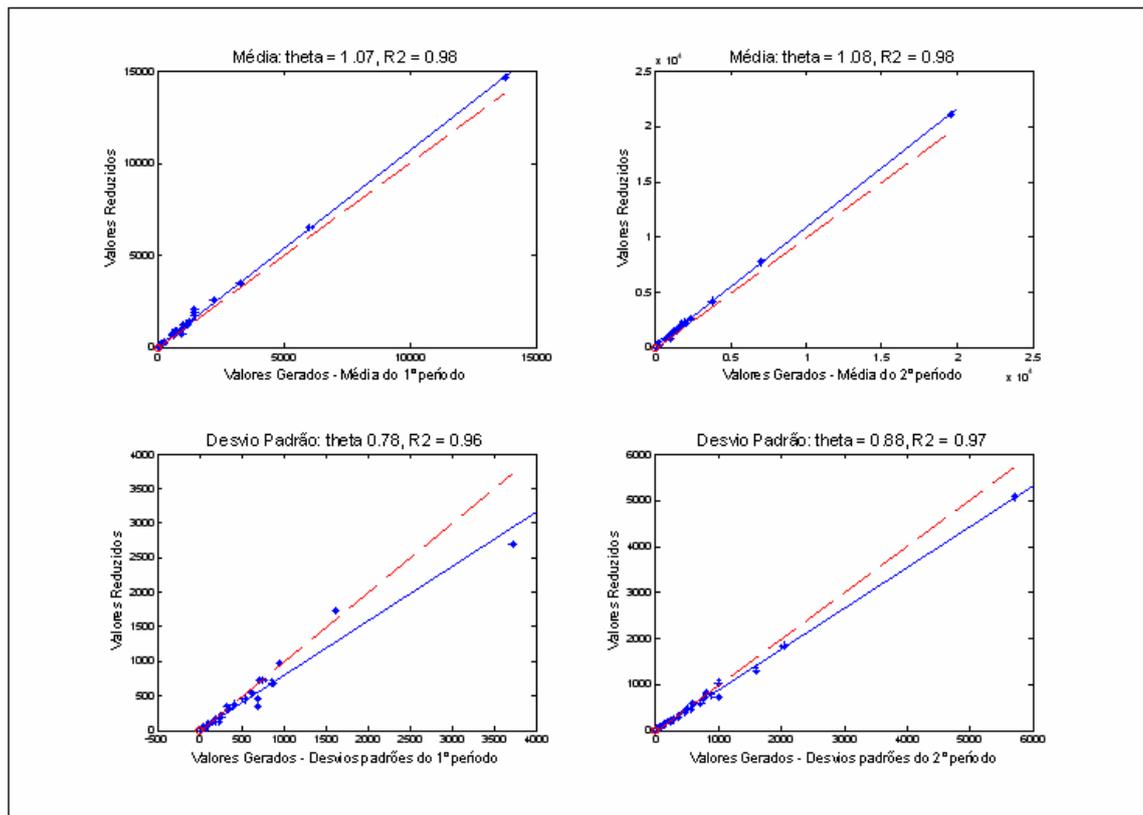


Figura 6.3: Regressão linear entre as médias e desvios padrões 1º e 2º períodos. Redução Local.

A Figura 6.3 apresenta bons resultados para as médias. Foram obtidos níveis aceitáveis para os resultados dos desvios padrões do 1º período; melhores resultados são obtidos para os desvios padrões do 2º período. Vale ressaltar que foram preservados apenas 1% do total de cenários gerados.

Para melhor analisar a redução local, apresenta-se a Figura 6.4 onde são representadas as regressões lineares de uma árvore de cenários construída a partir da agregação em conglomerados (esta possui a mesma estrutura da árvore construída pela redução local).

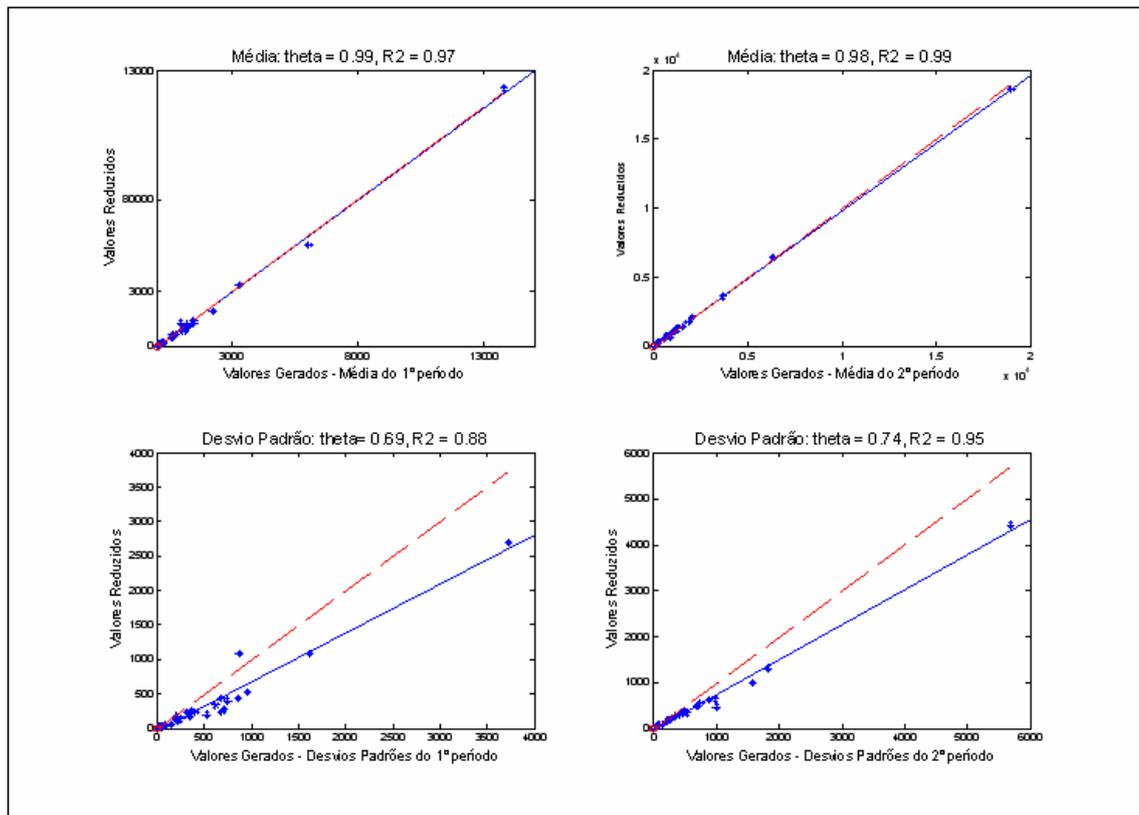


Figura 6.4: Regressão linear entre as médias e desvios padrões 1º e 2º períodos. Agregação.

A partir da Figura 6.4 pode-se verificar que, em se tratando de desvios padrões, a redução local é mais eficiente que a agregação em conglomerados.

Outras comparações entre as duas metodologias podem ser encontradas em [21].

7 - Conclusões

A elevada cardinalidade de informações necessárias para representar acuradamente a incerteza hidrológica no planejamento da operação energética de curto prazo torna o problema de otimização a ser resolvido inviável em termos de tempo computacional para o SIN. A abordagem por meio da ROC se apresenta estatisticamente adequada para selecionar um subgrupo, de cardinalidade fixa, com os cenários mais representativos do processo estocástico analisado.

O desenvolvimento teórico da ROC é fundamentado na seleção de eventos aleatórios de modo que se preserve a solução ótima e o valor funcional ótimo do problema de otimização estocástica subjacente. Como apresentado no Capítulo 4, a formulação da ROC é fundamentada na minimização da diferença entre os valores funcionais ótimos. Por esta razão, uma analogia entre o problema de selecionar cenários e o problema de transporte de Monge-Kantorovich se apresentou logicamente aplicável, quando se consideram como custo de transporte as probabilidades de ocorrência de cada cenário, e como distâncias entre as localidades de carga e descarga, as distâncias entre cenários. A relação de dualidade do problema de transporte, como apresentado no Capítulo 3, tornou favorável a satisfação do objetivo básico (selecionar cenários mantendo certos critérios de proximidade) por meio de fórmulas matemáticas compactas e de fácil aplicabilidade.

A aplicação da ROC no problema de planejamento da operação energética proporcionou a necessidade de se estudar formas diferenciadas do cálculo de distâncias entre cenários, devido à situação dos cenários de afluentes aos aproveitamentos hidroelétricos serem dispostos em formato árvore multivariada. Este assunto foi exposto no Capítulo 5, onde foram realizadas as análises de algoritmos e as apresentações de vários resultados do processo de ROC. Uma dificuldade apresentada pela aplicação da ROC neste contexto, é a necessidade de se armazenar um grande número de informações para calcular as distâncias entre os diferentes cenários. A proposta apresentada para contornar esta dificuldade é a utilização do algoritmo de redução local. Este algoritmo se apresenta eficiente tanto em termos de esforço computacional, quanto em termos de resultados obtidos.

Como pode-se verificar pelos resultados apresentados, a técnica de redução ótima de cenários é uma ferramenta eficiente para selecionar cenários suficientemente representativos, proporcionando boa aderência estatística entre o grupo e o subgrupo de cenários de vazões afluentes.

O uso da métrica de Fortet-Mourier para definir o problema de ROC e selecionar um subgrupo de cenários, a partir de uma árvore original com muitos cenários, permite eliminar redundâncias de informações, e reduzir significativamente o tempo de processamento computacional. Cabe mencionar que os tempos de CPU necessários para efetuar a ROC em uma árvore com 7500 cenários, reduzindo-a para 100 cenários, está na ordem de 50 minutos em um computador Pentium 4 de 3.00 GHz. O tempo de CPU necessário para se construir, por meio do algoritmo de redução local, uma árvore com 100 cenários obtidos a partir de 10000 cenários, é de aproximadamente 2 minutos.

A experiência numérica apresentada mostra uma margem percentual satisfatória nos

erros cometidos nas médias, nos desvios padrões, e nas correlações temporais e espaciais, quando aplicada a técnica de ROC.

O trabalho apresentado, se futuramente aplicado no modelo DECOMP para a operação de curto prazo do SIN, permitirá utilizar árvores de cenários de cardinalidade reduzida que representem adequadamente o processo estocástico das afluências e que sejam satisfatórias, tanto em termos de esforço computacional, quanto em termos de qualidade das soluções obtidas.

Apêndice A

Dualidade em Programação Linear

O objetivo deste apêndice é provar alguns fatos importantes relacionados a dualidade em programação linear.

Proposição 4 *Sejam $c, x \in \mathbb{R}^n, b \in \mathbb{R}^m$ e A uma matriz em $\mathbb{R}^m \times \mathbb{R}^n$. Considere o seguinte problema*

$$\begin{cases} \min_{x \in \mathbb{R}^n} & c'x \\ \text{s.a:} & Ax = b \\ & -x \leq 0, \end{cases} \quad (\text{A.1})$$

onde o vetor c representa os custos, sendo x a variável de decisão e b é um vetor associado à matriz de restrições A .

A representação dual do problema (A.1) é dada por:

$$\begin{cases} \max_{u \in \mathbb{R}^m} & b'u \\ \text{s.a:} & Au \leq c. \end{cases} \quad (\text{A.2})$$

Prova. Considere a função Lagrangiana do problema (\tilde{P})

$$L(x, u, v) = c'x - u'(Ax - b) - v'x$$

sendo $u \in \mathbb{R}^m$ e $v \in \mathbb{R}_+^n$ os multiplicadores de Lagrange. Resolver (\tilde{P}) é equivalente a resolver

$$\inf_x \sup_{u, v} L(x, u, v)$$

que pela dualidade, é o mesmo que:

$$\sup_{u, v} \inf_x L(x, u, v).$$

Calculando o \inf_x tem-se:

$$\begin{aligned} \inf_x L(x, u, v) &= \inf_x \{c'x - u'(Ax - b) - v'x\} \\ &= \inf_x \{c'x - u'Ax + u'b - v'x\} \\ &= \inf_x \{(c' - u'A - v')x + u'b\} \\ &= \inf_x \{(c - Au - v)'x + u'b\}. \end{aligned}$$

O conjunto em que vale $\inf_x L(x, u, v) > -\infty$ é chamado de conjunto viável, e é dado

por:

$$\{(u, v) \in \mathbb{R}^m \times \mathbb{R}_+^n : c - A'u = v\} \quad (\text{A.3})$$

Eliminando v em (A.3) e usando o fato de que $v \in \mathbb{R}_+^n$, passa-se a ter a seguinte desigualdade $A'u \leq c$. Logo,

$$\sup_{u,v} \inf_x L(x, u, v) = \sup_u \{u'b : A'u \leq c\},$$

e o dual do problema (\tilde{P}) é dado por

$$(D) \begin{cases} \max_{u \in \mathbb{R}^m} & b'u \\ \text{s.a:} & A'u \leq c, \end{cases}$$

como se queria mostrar. ■

Outra derivação do problema dual em PL pode ser obtida usando o Lema de Farkas.

Lema1 (Lema de Farkas) Para cada matriz $A \in \mathbb{R}^{m \times n}$ e cada vetor b em \mathbb{R}^n , o conjunto

$$\mathbf{S}_1 := \{x \in \mathbb{R}^n; Ax \leq 0 \text{ e } b'x > 0\}$$

é não vazio se, e somente se, o conjunto

$$\mathbf{S}_2 := \{u \in \mathbb{R}^m; A'u = b \text{ e } u \geq 0\}$$

é vazio.

Prova. (\Rightarrow) Suponha por absurdo que exista um par $(\bar{x}, \bar{u}) \in \mathbb{R}^n \times \mathbb{R}^m$, respectivamente, pontos viáveis de \mathbf{S}_1 e \mathbf{S}_2 . Neste caso tem-se que

$$0 < b'\bar{x} = \bar{u}'A\bar{x} \leq 0.$$

O que é claramente uma contradição. Logo, não pode existir o par (\bar{x}, \bar{u}) .

(\Leftarrow) Nesta parte é associada ao par $(\mathbf{S}_1, \mathbf{S}_2)$ um par primal/dual, de modo que a negação de um dos sistemas equivalha a inviabilidade de certo programa linear e ilimitação do seu dual.

Mais precisamente, considere o par primal/dual de programas lineares

$$(\tilde{P}) \begin{cases} \min_{x \in \mathbb{R}^n} & c'x \\ \text{s.a:} & Ax = b \\ & -x \leq 0 \end{cases} \text{ e } (D) \begin{cases} \max_{u \in \mathbb{R}^m} & b'u \\ \text{s.a:} & A'u \leq 0. \end{cases}$$

Suponha que \mathbf{S}_1 seja vazio, então (\tilde{P}) é inviável. Desta maneira,

$$\inf_{\emptyset} \sup_u \{0'x + u'(Ax - b) - v'x\} = \inf_{\emptyset} 0'x + \sup_{u,v} u'(Ax - b) - v'x = \infty.$$

Por outro lado, (D) é viável para $u = 0$. Resolver (D) é o mesmo que

$$\begin{aligned} \inf_u -b'u + \sup_x x'A'u &= \inf_u \sup_x -b'u + x'A'u \\ &= \inf_u \sup_x -b'u + (xA)'u \\ &= \inf_u \sup_x u'(Ax - b) \\ &= -\infty \text{ (pois } (Ax - b) \neq 0, x \text{ é inviável para } (\tilde{P})\text{)}, \end{aligned}$$

sendo x um multiplicador de Lagrange do problema (D). Deste modo, (D) é ilimitado e, portanto, deve existir uma direção de ilimitação \bar{u} , ou seja,

$$\exists \bar{u} \in \mathbb{R}^m : A'\bar{u} \leq 0 \text{ e } b'\bar{u} > 0.$$

Segue então que S_2 é não vazio. ■

Com o auxílio do Lema de Farkas, enuncia-se o critério de otimalidade em programas lineares:

Lema 2: (critério de otimalidade em programas lineares) O ponto $\bar{u} \in \mathbb{R}^m$ é uma solução do problema (A.2) se, e somente se, existe $\bar{x} \in \mathbb{R}^n$ tal que o par (\bar{x}, \bar{u}) satisfaça as relações seguintes

- i) $A'\bar{u} \leq c$; (viabilidade do dual)
- ii) $A\bar{x} = b$; (viabilidade do primal)
- iii) $\bar{x} \geq 0$; (viabilidade do primal)
- iv) $b'\bar{u} = c'\bar{x}$. (complementariedade)

Prova. (\Rightarrow) Seja \bar{u} uma solução do problema dual (\tilde{P}). Define-se os conjuntos de índices E e R e N correspondentes as restrições duais lineares ativas e inativas respectivamente:

$$E := \{i; A'_i\bar{u} = c_i\} \text{ e } R := \{i; A'_i\bar{u} < c_i\}; \text{ Com esta notação,}$$

$$\begin{aligned} v) \quad A'_E\bar{u} &= c_E; \\ vi) \quad A'_R\bar{u} &< c_R. \end{aligned}$$

Se $E = \emptyset$, existe que $\delta > 0$ tal que $A'\bar{u} - c \leq -\delta\mathbf{1}$, sendo $\mathbf{1} \in \mathbb{R}^n$ o vetor unitário. Deste modo, para cada $u \in \mathbb{R}^m$, pode-se encontrar um número real $\alpha > 0$ tal que

$$A'(\bar{u} + \alpha u) - c \leq -\delta\rho + \alpha A'u \leq 0,$$

e portanto $\bar{u} + \alpha u$ pertence ao conjunto viável de (\tilde{P}) . Como \bar{u} é solução de (\tilde{P}) , tem-se que para cada $u \in \mathbb{R}^m$ existe $\alpha > 0$ tal que

$$b'\bar{u} \geq b'(\bar{u} + \alpha u).$$

Portanto, $b'u \leq 0$ para cada $u \in \mathbb{R}^m$. Isto implica que $b = 0$ e, logo, tomando $\bar{x} = 0 \in \mathbb{R}^n$, as relações de *i)* a *iv)* são satisfeitas trivialmente.

Se $E \neq \emptyset$, então pode-se afirmar que o sistema

$$vii) \quad A'_E u \leq 0, \quad b'u > 0, \quad \text{não tem solução } u \in \mathbb{R}^m.$$

Suponha por contradição que $\exists \tilde{u} \in \mathbb{R}^m$ tal que as relações em *vii)* são satisfeitas. Então o vetor $\alpha\tilde{u}$ também satisfaz as relações, para qualquer $\alpha > 0$. O ponto $\bar{u} + \alpha\tilde{u}$, para $\alpha > 0$ arbitrário, satisfaz as desigualdades

$$\begin{aligned} viii) \quad b'(\bar{u} + \alpha u) &> b'\bar{u} && \text{(por } vii) \\ ix) \quad A'_E(\bar{u} + \alpha u) - c_E &\leq 0 && \text{(por } vii \text{ e } v) \\ x) \quad A'_R(\bar{u} + \alpha u) - c_R &\leq -\delta\rho + \alpha A'_R u \leq 0 && \text{(por } vi \text{ e } vii) \end{aligned}$$

A desigualdade a esquerda em *x)* conseqüência de definir δ como

$$-\delta := \max_{i \in R} (A'_i \bar{u} - c_i) < 0,$$

e a segunda desigualdade em *x)* vale para algum $\alpha > 0$, pelo fato de que $-\delta < 0$. Mas as relações de *vii)* a *x)* implicam que $\bar{u} + \alpha\tilde{u}$ é viável para (\tilde{P}) e que $b'(\bar{u} + \alpha u) > b'\bar{u}$. Chega-se assim numa contradição, já que \bar{u} é solução de (\tilde{P}) .

Portanto, *vii)* não tem solução $u \in \mathbb{R}^m$, e pelo Lema de Farkas, o sistema

$$xi) \quad A_E y = b, \quad y \geq 0$$

deve ter uma solução $y \in \mathbb{R}^e$, onde e é a cardinalidade do conjunto E . Seja r a cardinalidade de R e seja o vetor $0 \in \mathbb{R}^r$, tem-se que

$$\begin{aligned} xii) \quad A_E y + A_R 0 &= b, & y &\geq 0 \\ xiii) \quad c'_E y + c'_R 0 &= y' c_E = y' A'_E \bar{u} = b'\bar{u} && \text{(por } v \text{ e } xi). \end{aligned}$$

Definindo $\bar{x} \in \mathbb{R}^n$ tal que $\bar{x} = (\bar{x}_E, \bar{x}_R)$, sendo $\bar{x}_E := y \in \mathbb{R}^e$, $\bar{x}_R := 0 \in \mathbb{R}^r$, a condição *xiii*) se torna *iv*), a condição *xii*) se torna as condições *ii*) e *iii*). A condição *i*) vale porque \bar{u} pertence ao conjunto viável do problema (\tilde{P}).

Com isso, se prova que: se \bar{u} é solução ótima de (\tilde{P}), existe \bar{x} tal que as relações *i*) a *iv*) são satisfeitas, como se queria mostrar.

(\Leftarrow) Sejam $\bar{u} \in \mathbb{R}^m$ e $\bar{x} \in \mathbb{R}^n$ satisfazendo as condições *i*) a *iv*). Para u , um ponto qualquer pertencente ao conjunto viável do problema (\tilde{P}), as relações a seguir resultam de simples manipulações algébricas:

$$\begin{aligned} b'u - b'\bar{u} &= (A\bar{x})'u - b'\bar{u} \quad (\text{por } ii) \\ &= \bar{x}'A'u - c'\bar{x} \quad (\text{por } iv) \\ &= \bar{x}'(A'u - c') \\ &\leq 0. \quad (\text{por } iii, v \text{ e } vi). \end{aligned}$$

Logo $b'\bar{u} \geq b'u$ para qualquer u viável. Portanto \bar{u} é solução do problema (\tilde{P}). ■

Apêndice B

Relações entre os Multiplicadores de Lagrange do Problema de Monge-Kantorovich Caso Discreto

A seguir será provada a proposição apresentada no Capítulo 3, página 27.

Proposição 2: *Considere o caso discreto em que o espaço amostral Θ_P contém o espaço amostral Θ_Q , isto é,*

$$\Theta_P = \{w_1, \dots, w_N\} \text{ e } \Theta_Q = \{\tilde{w}_1, \dots, \tilde{w}_M\} = \{w_{j_1}, \dots, w_{j_M}\},$$

sendo $\{w_{j_1}, \dots, w_{j_M}\} \subset \Theta_P$. Considere também a função μ satisfazendo as condições i) e ii) de função distância da página 27. Com esta convenção o par ótimo (\bar{u}, \bar{v}) , associado ao problema (3.9), possui a seguinte relação:

$$\bar{u}_i = -\bar{v}_{j_k}, \text{ se } i = j_k.$$

Prova. A demonstração é feita utilizando a representação primal do problema (3.9). Pelo resultado de dualidade do problema MK, tem-se que

$$\begin{aligned} \gamma_\mu(P, Q) &:= \max_{u, v} \left\{ \sum_{i=1}^N u_i p_i + \sum_{k=1}^M v_{j_k} q_{j_k} : u_i + v_{j_k} \leq \mu(w_i, w_{j_k}) \right\} \\ &= MK(P, Q), \end{aligned}$$

sendo

$$MK(P, Q) := \min_{\eta} \left\{ \sum_{i=1}^N \sum_{k=1}^M \mu(w_i, w_{j_k}) \eta_{ij_k} : \eta_{ij_k} \geq 0, \sum_{i=1}^N \eta_{ij_k} = q_{j_k}, \sum_{k=1}^M \eta_{ij_k} = p_i \right\}$$

o problema primal.

Afirma-se que, se $\bar{\eta}$ é solução do problema primal, então

$$s = j_l \Rightarrow \bar{\eta}_{sj_l} > 0.$$

Para verificar a afirmação, suponha que $\bar{\eta}_{sj_l} = 0$. Sendo $s = j_l$, tem-se que $\mu(w_s, w_{j_l}) = 0$; logo $\bar{\eta}_{sj_l}$ não influencia no valor funcional do problema $MK(P, Q)$. Aumentando o valor $\bar{\eta}_{sj_l}$ para $\tilde{\eta}_{sj_l} > 0$, a restrição $\sum_{k=1}^M \bar{\eta}_{sj_k} = p_s$ implica em uma redução do valor $\bar{\eta}_{sj_r}$ ($\bar{\eta}_{sj_r} := \tilde{\eta}_{sj_r}$) para algum índice $r < M$. Devido à restrição $\sum_{i=1}^N \bar{\eta}_{ij_r} = q_{j_r}$, tem-se um crescimento no valor $\bar{\eta}_{ij_r}$ ($\bar{\eta}_{ij_r} := \tilde{\eta}_{ij_r}$) para algum $t < N$; pode-se então escolher

$t = j_r$ para que o valor da função objetivo não aumente ($t = j_r \Rightarrow \mu(w_t, w_{j_r}) = 0$). De forma análoga, a restrição $\sum_{k=1}^M \bar{\eta}_{j_r j_k} = p_{j_r}$ implica em uma redução do valor $\bar{\eta}_{j_r j_n}$ ($\bar{\eta}_{j_r j_n} := \tilde{\eta}_{j_r j_n}$) para algum $n < M$; que por sua vez implica em um crescimento do valor $\bar{\eta}_{e j_n}$. Basta então escolher $n = l$ e $e = s$, pois $\bar{\eta}_{s j_l}$ ($\bar{\eta}_{s j_l} := \tilde{\eta}_{s j_l}$) foi inicialmente acrescido. Em resumo, tem-se que:

$\bar{\eta}_{s j_l}$ aumentou; como $s = j_l$, segue que o valor funcional não se alterou;

$\bar{\eta}_{s j_r}$ diminuiu; sendo $s \neq j_r$, segue que o valor funcional diminuiu;

$\bar{\eta}_{t j_r}$ aumentou; como $t = j_r$, segue que o valor funcional não se alterou;

$\bar{\eta}_{j_r j_n}$ diminuiu; sendo $j_r \neq j_n$, segue que o valor funcional diminuiu;

Reorganizando os índices, $j_n = j_l = s$ e $t = j_r$ segue que o par $(\bar{\eta}_{st}, \bar{\eta}_{ts})$ teve uma redução em seu valores, e o par $(\bar{\eta}_{ss}, \bar{\eta}_{tt})$ aumentou seu valores.

Definindo η^* como

$$\eta_{il}^* = \begin{cases} \bar{\eta}_{il}, & \text{se } i \neq s \text{ e } l \neq t \\ \tilde{\eta}_{il}, & \text{caso contrário,} \end{cases}$$

tem-se que η^* é viável e proporciona um valor funcional menor para o problema primal; contrariando assim a hipótese inicial de que $\bar{\eta}$ é solução.

Logo, $s = j_l \Rightarrow \bar{\eta}_{s j_l} > 0$ provando assim a afirmação.

Passa-se agora à prova da proposição. Utilizando a notação do Capítulo 3 (página 21), pode-se escrever o problema primal como

$$MK(P, Q) = \begin{cases} \min_X & \bar{C}' X \\ \text{s.a.} & AX = b \\ & -X \leq 0. \end{cases}$$

Escrevendo a função lagrangiana deste problema, tem-se que

$$L(X, \Lambda, \Pi) = \bar{C}' X + \Lambda'(AX - b) - \Pi' X,$$

sendo $\Lambda \in \mathbf{R}^{N+M}$ e $\Pi \in \mathbf{R}_+^{N+M}$ os multiplicadores de Lagrange.

Se a variável \bar{X} é solução do problema $MK(P, Q)$, então existem $(\bar{\Lambda}, \bar{\Pi})$ tais que satisfazem o sistema de otimalidade de Lagrange ($\nabla L(\bar{X}, \bar{\Lambda}, \bar{\Pi}) = 0$, ver [12]). Ou seja,

$$\begin{aligned} i) \quad & A' \bar{\Lambda} = \bar{\Pi} - \bar{C}; \\ ii) \quad & A \bar{X} = b; \\ iii) \quad & \bar{\Pi}' \bar{X} = 0. \end{aligned}$$

A igualdade em i) equivale a

$$\begin{cases} \bar{u}_1 + \bar{v}_{j_1} & = \bar{\Pi}_1 - \mu(w_1, w_{j_1}) \\ & \vdots \\ \bar{u}_1 + \bar{v}_{j_M} & = \bar{\Pi}_M - \mu(w_1, w_{j_M}) \\ \bar{u}_2 + \bar{v}_{j_1} & = \bar{\Pi}_{M+1} - \mu(w_2, w_{j_1}) \\ & \vdots \\ \bar{u}_N + \bar{v}_{j_M} & = \bar{\Pi}_{NM} - \mu(w_N, w_{j_M}). \end{cases}$$

Suponha que $s = j_l$, logo $\bar{u}_s + \bar{v}_{j_l} = \bar{\Pi}_{M(s-1)+l} - 0$. Como já provado, $s = j_l \Rightarrow \bar{\eta}_{sj_l} > 0$ que equivale a $\bar{X}_{M(s-1)+l} > 0$. Utilizando a igualdade iii), segue que $0 = \bar{\Pi}_{M(s-1)+l} \Rightarrow \bar{u}_s = -\bar{v}_{j_l}$.

Com este resultado, basta verificar que o par $(\bar{u}, \bar{v}) = \bar{\Lambda}$ é solução do problema dual $\gamma_\mu(P, Q)$. Sendo \bar{X} solução do problema primal, pelo resultado de dualidade fraca,

$$b' \bar{\Lambda} \leq C' \bar{X} \text{ para todo } \bar{\Lambda} \text{ viável,}$$

e utilizando o Lema de Farkas (2), tem-se que

$$\begin{aligned} A' \bar{\Lambda} &\leq \bar{C}; \\ A \bar{X} &= b; \\ \bar{X} &\geq 0; \\ b' \bar{\Lambda} &= \bar{C}' \bar{X}. \end{aligned}$$

Logo $\bar{\Lambda}$ é solução do problema primal $\gamma_\mu(P, Q)$ como se queria mostrar. ■

Referências Bibliográficas

- [1] BONNANS, J.F.; GILBERT, J.C.; LEMARÉCHAL, C.; SAGASTIZÁBAL, C. (2002). " *Numerical Optimization*" Theoretical and Practical Aspect. Springer-Verlag, Heidelberg. pp 423.
- [2] BOX, G.E.P.; JENKINS, G.M. (1994) " *Time Series Analysis, Forecasting and Control*." San Francisco; Holden-Day. (Third Edition).
- [3] CENTRO DE PESQUISAS DE ENERGIA ELÉTRICA. (2002). Projeto Gevazp, " *Modelo de Geração de Séries Sintéticas de Energia e Vazões - Manual de Referência*" . Cepel, 2002. pp. 64.
- [4] COSTA, J.P.; OLIVEIRA, G.C.; LEGEY, L.F.L. (2006). " *Reduced Scenario Tree Generation for Mid-Term Hydrothermal Operation planning*" . 9th International Conference in Probabilistic Methods Applied to Power Systems KTH, Stockholm, Sweden. pp. 11-15.
- [5] DAMÁZIO, J.M. (2002). " *Verificação da preservação dos dois primeiros momentos da distribuição condicionada mais de um passo-à-frente em séries geradas em árvore pelo modelo GEVAZP*" . Relatório Técnico, CEPEL, pp. 8-14.
- [6] DUPAČOVÁ, J.; GRÖWE-KUSKA, N.; RÖMISCH, W. (2003). " *Scenario reduction in stochastic programming: An approach using probability metrics*" Mathematical Programming, Ser. A 95, 493-511.
- [7] FORTUNATO, L.M.; NETO, T.A.A.; ALBUQUERQUE, J.C.R.; PEREIRA, M.V.F. (1990). " *Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica*" . Universidade Federal Fluminense. EDUFF- Editora Universitária.
- [8] GRÖWE-KUSKA, N.; HEITSCH, H.; RÖMISCH, W. (2003). " *Scenario reduction and scenario tree construction for power management problems*" , IEEE Bologna Power Tech Proceedings (A. Borghetti, C.A. Nucci, M. Paolone eds.), IEEE.
- [9] GUJARATI, D.N.(2000). " *Econometria Básica*" . Makron Books, 3th ed. pp. 64-70.
- [10] HEITSCH, H.; RÖMISCH, W. (2003). " *Scenario reduction algorithms in stochastic programming*" , Computational Optimization and Applications 24 (2003), 187-206.

- [11] HOCHREITER, R.; PFLUG, G. (2002). " *Scenario Tree Generation as a Multidimensional Facility Location Problem* ", Aurora TR2002-33, pp. 5-8.
- [12] IZMAILOV, A.F.; SOLODOV, M.V. (2005) " *Otimização, vol.1: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade* ". Rio de Janeiro: IMPA, pp. 219-236.
- [13] JARDIM, D.L.D.D.; MACEIRA, M.E.P.; FALCÃO, D.M. (2001). " *Stochastic Streamflow Model for Hydroelectric Systems Using Clustering Techniques*" In Anais do IEEE Porto Power Tech Conference, vol.3, Porto, Portugal, Sep., pp 3-5.
- [14] KELMAN, J. (1987). " *Modelos Estocásticos no Gerenciamento dos Recursos Hídricos* ". In : Modelos para Gerenciamento de Recursos Hídricos, Coleção ABRH de Recursos, São Paulo.
- [15] KELMAN, J.; OLIVEIRA, G.C.; MACEIRA, M.E.P. (1992). " *Modelo de Séries Hidrológicas (MSH) -- Manual de Metodologia* ", Relatório Técnico CEPEL nº 242/92.
- [16] MACEIRA, M.E.P.; BEZERRA, C.V. (1997). " *Stochastic Streamflow model for Hydroelectric Systems*" In: Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems, Vancouver, Canada, Sep., pp. 305-310.
- [17] MACEIRA, M.E.P.; JARDIM, D.L.D.D. (2000). " *Modelo de Geração de Séries Sintéticas de Energia e Vazões -- GEVAZP -- Manual de Referência*" In: Relatório Técnico DPP/PEL 83/2000, CEPEL, Rio de Janeiro.
- [18] MACEIRA, M.E.P.; PENNA, D.D.J.; DAMÁZIO, J.M. (2005). " *Geração de Cenários Sintéticos de Energia e Vazão para o Planejamento da Operação Energética*" In: XVI Simpósio Brasileiro de Recursos Hídricos, João Pessoa, Paraíba, Brasil, pp. 3-12.
- [19] MACEIRA, M.E.P.; TERRY, L.A.; DAMAZIO, J.M.; COSTA, F.S., MELO, A.C.G. (2002). " *Chain of Models for Setting the Energy Dispatch and Spot Price in the Brazilian System, Power*" System Computation Conference -- PSCC'02, Sevilla, Spain, June 24-28, pp. 1-4.
- [20] MANGASARIAN, O.L. (1969). " *Nonlinear Programming* ". McGraw-Hill, New York.
- [21] OLIVEIRA, W.L.; SAGASTIZÁBAL, C.; PENNA, D.D.J.; MACEIRA, M.E.P.; DAMÁZIO, J.M. (2007). " *Redução ótima de Cenários de Vazões Afluentes* "

aos Reservatórios" Aceito em: XVII Simpósio Brasileiro de Recursos Hídricos.

[22] RASCHEV, S.T.; RÜSCHENDORF, L. (1998) " *Mass Transportation Problems*" Volume I: Theory (Probability and its Applications) Springer; 1st. ed., pp. 1-8 e 57-76.

[23] RASMUSSEM, K.M.; CLAUSEN, J. (2007). " *Mortgage loan portfolio optimization using multi-stage stochastic programming*" Journal of Economic Dynamics and Control, Volume 31, Issue 3, pp. 742-766.

[24] RUSZCZYŃSKI, A. (1997). " *Decomposition Methods in Stochastic Programming*" , Mathematical Programming, 79; pp. 333-353.

[25] SALAS, J.D; DELLEUR, J.W.; YEVIEVICH, V.; LANE W.L. (1980). " *Applied Modeling of Hydrologic Time Series*" , Water Resources Publications, p. 484.

[26] SEIGEL, S. (1975) " *Estatística Não Paramétrica*" , São Paulo, Mc Graw-Hill. pp. 144-155.

[27] XAVIER, L.N., DINIZ, A.L.S., COSTA, F.S., MACEIRA, M.E.P (2005). " *Aprimoramento da Modelagem da Função de Produção Energética das Usinas Hidroelétricas no Modelo DECOMP: Metodologias e Resultados*" XVIII SNPTEE - - Seminário Nacional de Produção e Transmissão de Energia Elétrica, Curitiba, PR.

[28] OLIVEIRA, W.L. (2007). "Redução ótima de cenários em programação estocástica. Aplicação às vazões afluentes aos aproveitamentos hidroelétricos" Dissertação de Mestrado - IMPA.