Fabio Gagliardi Cozman

Tomada de Decisão com Conjuntos de Medidas de Probabilidade:

Conceitos de Independência e Modelos Baseados em Grafos

Texto sistematizado apresentado à Escola Politécnica da Universidade de São Paulo para o Concurso de Livre Docência junto ao Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos, na especialidade Automação da Manufatura e Robótica.

São Paulo

2003

Sumário

1 Introdução **6**

1.1 Experiência do autor 7

1.2 Temas de pesquisa no período 1998-2002 10

1.2.1 Breve sumário de resultados obtidos na teoria de conjuntos de probabilidade 10

1.2.2 Breve descrição da pesquisa em redes Bayesianas 11

1.2.3 Breve descrição da pesquisa em automação industrial e robótica 13

1.2.4 Breve descrição da pesquisa em dispositivos para automatização da marcha de pessoas com deficiências motoras 14

2 A teoria de conjuntos credais **15**

2.1 Conjuntos credais em inteligência artificial 15

2.2 Conceitos básicos 17

2.3 Valores esperados condicionais inferiores 19

2.4 Verossimilhança e valores esperados condicionais inferiores 22

2.5 Condicionando com probabilidade inferior nula 24

2.6 Discussão 26

3 Independência para conjuntos credais **27**

3.1 Conceitos de independência 28

3.1.1 Independência epistêmica 28

3.1.2 Independência forte 29

3.1.3 Condição de Kuznetsov 30

3.2 Independência e permutabilidade 30

3.3 Independência forte e semi-grafóides 33

3.4 Independência epistêmica e semi-grafóides 34

3.5 Condição de Kuznetsov e semi-grafóides 36

3.6 Discussão 39

4 Redes credais e a condição de separação **41**

4.1 Modelos baseados em grafos direcionados acíclicos 42

4.1.1 Redes Bayesianas 42

4.1.2 Redes credais 43

4.2 Extensões fortes 44

4.3 Extensões epistêmicas 45

4.4 A condição de separação 47

4.5 Discussão 48

5 Conclusão **49**

Lista de Figuras

1.1 Linhas e projetos de pesquisa. 9

4.1 Exemplo de rede Bayesiana. 42

4.2 Exemplo de rede credal com quatro variáveis binárias. 43

4.3 Exemplo de rede credal: todas as variáveis são binárias, exceto C, que tem três valores. 44

Lista de Tabelas

3.1 Densidades que definem conjunto credal no Exemplo 4. 36

3.2 Vértices do conjunto credal no Exemplo 5. 39

Capítulo 1

Introdução

Este documento sistematiza atividades de pesquisa do autor, focando sobretudo no período compreendido entre janeiro de 1998 e dezembro de 2002. O autor terminou seu curso de doutoramento em dezembro de 1997 na Carnegie Mellon University, Estados Unidos, e imediatamente a seguir iniciou suas atividades de docência e pesquisa na Universidade de São Paulo.

Neste capítulo introdutório estão descritas as atividades de pesquisa do autor de forma geral, para permitir uma visão panorâmica de resultados obtidos em vários temas e projetos. Os demais capítulos se concentram em uma particular linha de pesquisa desenvolvida pelo autor, procurando sistematizar um conjunto articulado de publicações. O foco desse esforço de sistematização é a *teoria de conjuntos de medidas de probabilidade*. O objetivo global do trabalho descrito é construir uma teoria de conjuntos de medidas de probabilidade que inclua, de forma organizada, conceitos de independência entre variáveis.

A teoria básica sobre conjuntos de medidas de probabilidade é apresentada no Capítulo 2. Vários conceitos de independência são analisados no Capítulo 3. A partir dessa análise, o Capítulo 4 discute como construir modelos multivariados a partir de grafos e conjuntos de medidas de probabilidade. A combinação de conceitos de independência e modelos multivariados culmina em uma condição de separação que relaciona os principais conceitos da teoria. Para facilitar o entendimento, o material foi colocado de forma contínua e coerente, sendo feitas referências às publicações do autor que apresentam provas e detalhes técnicos.

1.1 Experiência do autor

O autor iniciou sua carreira acadêmica como estagiário em 1988, no grupo de Mecatrônica da Escola Politécnica, então lotado no Departamento de Engenharia Mecânica da Escola Politécnica. O autor recebeu o título de Mestre em Engenharia (aprovado com distinção e louvor) pela Escola Politécnica em dezembro de 1991. Após a conclusão do mestrado, trabalhou como pesquisador visitante no Center for Systems Science, Yale University (Estados Unidos) por um mês. Iniciou então programa de doutoramento no Robotics Institute, Carnegie Mellon University (Pittsburgh, Estados Unidos), onde recebeu o título de PhD in Robotics em dezembro de 1997.

Como estagiário e mestrando, o autor trabalhou no sistema de controle de um robô móvel (termo que designa veículos sem motorista humano e com capacidade de trafegar por rotas livremente escolhidas). O trabalho se concentrou nos níveis de controle relacionados a manutenção de trajetória. Controladores foram projetados segundo a teoria de controle ótimo, tanto para modelos dinâmicos baseados em referenciais fixos quanto para referenciais móveis [3, 30]. Após o término do mestrado, o autor fez um estágio de um mês no Center for Systems Science, Yale University, Estados Unidos, centrado em técnicas de estimação para robôs móveis. Esse trabalho foi seguido do programa de doutoramento na Carnegie Mellon University.

O trabalho de doutorado teve duas fases: na primeira, caracterizada por cursos e exames qualificatórios, houve participação em projetos de robótica espacial. Foram realizados experimentos com o sistema de controle e planejamento do robô móvel Ratler e com métodos de tele-operação [104]. Durante o trabalho, foram também feitos experimentos com sistemas de localização baseados em informação celestial [22] e baseados em medidas visuais de condições atmosféricas [25]. A segunda fase do doutorado foi caracterizada pela concentração em um tema único de pesquisa. O trabalho defendeu a tese que conjuntos convexos de distribuições de probabilidade formam uma poderosa ferramenta para atividades de tomada de decisão em robótica e inteligência artificial. A principal contribuição de trabalho foi o desenvolvimento de algoritmos para inferência e estimação em duas aplicações distintas. A primeira aplicação foi a análise de robustez em modelos gráficos de inferência. A segunda aplicação explorada no trabalho foi a estimação de posição para robôs móveis em ambientes externos usando estimativas de probabilidade imprecisas. Para realizar experimentos com os algoritmos desenvolvidos

no trabalho, foi implementado um sistema de manipulação de redes Bayesianas denominado JavaBayes. Esse sistema têm sido distribuído via internet e têm sido utilizado em universidades e centros de pesquisa ao redor do mundo. Já a pesquisa em robótica produziu o sistema Viper, que foi utilizado com sucesso no robô Nomad em missão de 45 dias no deserto do Atacama (Chile). Além dos sistemas JavaBayes e Viper, a pesquisa resultou em várias publicações [23, 29, 24, 26, 66, 28, 27, 47, 34] (note-se que algumas dessas publicações vieram a ocorrer no período 1998-2000, após refinamentos e revisões).

Após o doutoramento, o autor assumiu a posição de Professor Doutor na Escola Politécnica da Universidade de São Paulo. As pesquisas no período 1998-2002 podem ser divididas em alguns temas, conforme ilustrado na Figura 1.1:

- Dentro da linha de manipulação de modelos estatísticos:
  - Algoritmos para manipulação e aplicações de conjuntos de probabilidades.
  - Algoritmos para utilização de redes Bayesianas em sistemas embarcados, em sistemas para classificação de dados e em sistemas de suporte à decisão médica.
- Dentro da linha de dispositivos robóticos:
  - Desenvolvimento de sistema de visão computacional para robô móvel de pequeno porte.
  - Automatização do processo de marcha de pessoas com deficiências motoras.

Estes temas de pesquisa estão discutidos na Seção 1.2. O seguinte resumo quantitativo pode ser feito do trabalho no período 1998-2002:

- 2 livros (anais de eventos) editados;
- 5 publicações em periódicos internacionais (uma delas convidada sem revisão);
- 1 publicação em periódico nacional;
- 21 publicações em eventos internacionais e nacionais com seletiva política editorial;

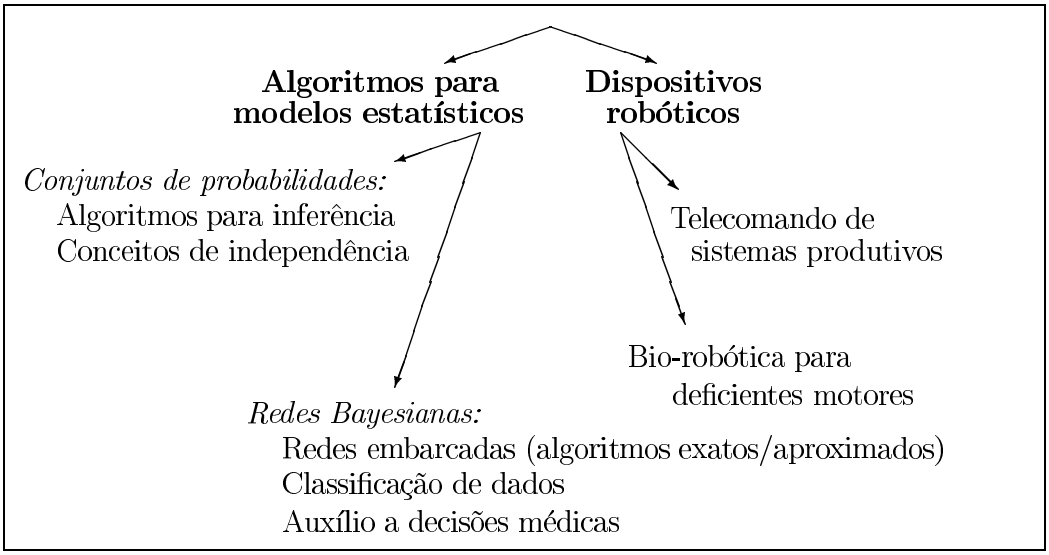


Figura 1.1: Linhas e projetos de pesquisa.

- 12 publicações completas ou resumos em outros eventos internacionais e nacionais;
- 8 participações na organização de eventos internacionais e nacionais;
- 4 participações na editoração de periódicos internacionais.

Além disso, foram conduzidas atividades regulares de pareceirista da FAPESP e CNPq, revisor de artigos de conferências e periódicos nacionais e internacionais, bem como atividades de orientação de mestrado e doutorado (dois mestrados concluídos), atividades de ensino e coordenação de graduação na Universidade de São Paulo, e atividades de extensão tanto no curso PECE como através de convênio com o Instituto de Pesquisas Eldorado. As atividades de coordenação de graduação relacionadas ao curso de Engenharia Mecatrônica levaram a duas publicações no Congresso Nacional de Ensino de Engenharia [4, 46]. Todas essas atividades não relacionadas diretamente à pesquisa não são abordadas no restante desta dissertação.

1.2 Temas de pesquisa no período 1998-2002

Esta seção procura fornecer resumidamente os principais resultados obtidos nos tópicos apresentados na Figura 1.1. Como a teoria de conjuntos de probabilidade é o objeto principal deste trabalho e será discutida em detalhes adiante, a Seção 1.2.1 apenas descreve os resultados e publicações principais. Os demais tópicos de pesquisa (redes Bayesianas, robô móvel e dispositivos para deficientes) são brevemente descritos nas Seções 1.2.2, 1.2.3 e 1.2.4.

1.2.1 Breve sumário de resultados obtidos na teoria de conjuntos de probabilidade

A pesquisa relacionada à teoria de conjuntos de probabilidade foi centrada em algoritmos e conceitos de independência. Foram abordados algoritmos para modelos baseados em grafos, para situações onde observações tem probabilidade zero, para experimentos repetitivos, e particularmente algoritmos que manipulem relações de independência entre variáveis. Resultados teóricos foram publicados em vários periódicos e conferências no período 1998-2002 [28, 27, 32, 53, 36, 31, 35, 33, 34, 41, 48, 42, 50, 49, 45], nos três últimos casos com co-autoria dos alunos de doutorado José Carlos Ferreira da Rocha e Cassio Polpo de Campos. Os resultados contidos nestas publicações estão descritos em capítulos posteriores desse documento. Deve-se notar que o artigo *Independence Relations in the Robustness Analysis of Multivariate Probabilistic Models*, apresentado na XII Conferência Brasileira de Automática, foi um dos seis artigos selecionados pelo comitê organizador para publicação na Revista Ciência e Engenharia.

Além das publicações mencionadas, o período 1998-2002 envolveu uma série de atividades profissionais relevantes.

Em 1999, o autor participou como co-organizador do primeiro *International Symposium on Imprecise Probabilities and Their Applications* (realizado em Ghent, Bélgica) e co-editou os anais deste evento. O simpósio foi o primeiro encontro internacional de pesquisadores envolvidos com imprecisões em valores de probabilidade, e atraiu trabalhos de todo o mundo. Como resultado da co-organização do *First Symposium on Imprecise Probabilities and Their Applications*, o autor foi convidado a editar números

especiais dos periódicos *International Journal of Approximate Reasoning* e *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. Os números especiais, publicados em 2000, contêm artigos selecionados do primeiro *International Symposium on Imprecise Probabilities and Their Applications*. Devido ao sucesso deste primeiro simpósio, um segundo simpósio com o mesmo tema foi realizado em 2001 (em Ithaca, Estados Unidos), e o autor participou como membro do Steering Committee. Um terceiro simpósio deverá ser realizado em 2003 (em Lugano, Suiça).

Duas atividades profissionais foram desenvolvidas com o objetivo de fortalecer os laços entre pesquisadores relacionados à teoria de conjuntos de probabilidade:

- Entre 1999 e 2001, o autor foi co-editor do *Imprecise Probabilities Project*, um trabalho de coleta de informações sobre conjuntos de probabilidades iniciado pelos pesquisadores Gert de Cooman e Peter Walley.
- O autor foi escolhido editor do periódico informativo da *Society for Imprecise Probability Theory and Applications (SIPTA)* que teve seu primeiro número em janeiro de 2003. A sociedade foi criada em 2002 para congregar pesquisadores com interesse em conjuntos de probabilidades.

1.2.2 Breve descrição da pesquisa em redes Bayesianas

Uma rede Bayesiana é uma representação para um modelo estatístico multivariado, composta por um grafo direcionado acíclico e várias distribuições de probabilidade. Este modelo têm ganho enorme popularidade nos últimos dez anos, tanto no campo de inteligência artificial como em áreas de estatística e economia. A definição precisa e referências podem ser encontradas na Seção 4.1.1; nesta seção são apenas descritos os tópicos de pesquisa abordados pelo autor no período 1998-2002.

O autor iniciou sua pesquisa em redes Bayesianas já no período de doutoramento. Um dos resultados desse trabalho foi o sistema JavaBayes, que permite a manipulação de redes Bayesianas e é distribuído livremente através da internet. O sistema JavaBayes foi continuamente aprimorado no período 1998-2002, e é hoje utilizado ao redor do mundo em universidades e companhias para ensino e pesquisa. O sistema é brevemente descrito em publicação convidada no periódico editado pela *International Society for Bayesian Analysis* [39], e os algoritmos implementados no sistema JavaBayes

estão descritos em publicação nacional [38]. Como especialista em modelos estatísticos gráficos, foi membro convidado de painel na *XII Conference on Uncertainty in Artificial Intelligence* em 1998 e foi convidado para proferir palestra no *Congresso Nacional de Automação Industrial* em 1998. Também 1998 o autor foi membro convidado em comitê para normatização de formatos de redes Bayesianas (comitê organizado pela *Association for Uncertainty in Artificial Intelligence*), em virtude de sua experiência com o sistema JavaBayes. Outro evento notável do ano de 1998 foi a escolha do artigo *JavaBayes: Tomada de Decisões através de Modelos Estatísticos Gráficos* como um dos 3 melhores trabalhos em Informática no Prêmio SUCESU-SP 1998 de Tecnologia da Informação.

Em termos de pesquisa em redes Bayesianas, o período 1998-2002 procurou se concentrar no desenvolvimento de algoritmos para sistemas embarcados, na construção de classificadores a partir de bases de dados, e na construção de sistemas de suporte à decisão médica.

Considerável esforço foi dispendido na concepção de algoritmos para redes Bayesianas em sistemas com baixa disponibilidade de memória. A motivação para este tipo de algoritmo é a crescente importância de sistemas embarcados em veículos e equipamentos industriais e domésticos, em virtude da miniaturização de componentes eletrônicos. O desafio é atender aos requerimentos de memória destes sistemas. O trabalho evoluiu para uma publicação preliminar sobre aplicações de redes Bayesianas juntamente a controladores programáveis industriais [95], em parceria com o orientado Fabio Tozeto Ramos, e para publicações internacionais [93] e nacionais [94]. Na mesma linha voltada a sistemas embarcados, o orientado Jaime Shinsuke Ide desenvolveu algoritmos baseados em método Monte Carlo para inferência [72] e para testes [73, 74, 75]. Foram também abordados pelo autor métodos para redução de consumo de memória em algoritmos de estimação de redes Bayesianas [40].

Durante o período 2000-2002, foi consolidada uma parceria entre o autor e pesquisadores do Hewlett-Packard Laboratories (HP Labs) de Palo Alto, Estados Unidos, em temas relacionados a redes Bayesianas. A partir de contatos e discussões técnicas em 2000, o autor realizou uma visita de dois meses em 2001 ao HP Labs. Essa visita estabeleceu uma intensa colaboração centrada na utilização de redes Bayesianas para classificação de dados. Durante a visita foram produzidos os relatórios técnicos [43] e [17], que evoluíram para publicação internacional [44]. Foi também estabelecido

um convênio em 2002 entre a Universidade de São Paulo e o Instituto de Pesquisas Eldorado, este último como representante de pesquisadores do HP Labs. O autor têm sido coordenador deste convênio desde seu início, em um trabalho que objetiva o desenvolvimento de algoritmos para construção de redes Bayesianas a partir de bases de dados. O trabalho está em andamento, com algumas publicações preliminares [5, 6].

Finalmente, a pesquisa em redes Bayesianas também têm envolvido, no período 2001-2002, aplicações na área médica. O projeto *Desenvolvimento de Ferramentas Computacionais para Suporte à Tomada de Decisão Médica para Saúde Pública*, com financiamento da FAPESP, tem como objetivo a construção de sistemas de suporte para avaliação de pacientes com dores cardíacas. O projeto surgiu de contatos com médicos do Hospital Universitário da Universidade de São Paulo, em particular o Dr. Paulo Lotufo, e contou com a colaboração fundamental da Dra. Veerle Coupé pesquisadora visitante da Holanda com ampla experiência em tomada de decisões médicas. Um dos objetivos do projeto é investigar ferramentas para análise de sensibilidade das redes Bayesianas desenvolvidas, utilizando técnicas consagradas [19, 80] ou estendendo o trabalho do autor nesta direção [37]. Um funcionário da Universidade de São Paulo, André Hideaki Saheki, participa deste projeto.

O objetivo global dessas atividades é criar um grupo de pesquisa bastante forte, com reconhecimento internacional, em modelos estatísticos baseados em redes Bayesianas.

1.2.3 Breve descrição da pesquisa em automação industrial e robótica

Durante o período 1998-2002, o autor desenvolveu duas atividades em automação e robótica, que capitalizaram em sua experiência passada.

- O autor participou como co-responsável técnico do projeto *Tele-comando e Monitoração Remota de Sistemas Produtivos*, entre 1999 e 2001. Este projeto, por sua vez, era parte de uma iniciativa de pesquisa de grande porte financiado pela FAPESP e coordenada pelo Prof. Paulo Eigi Miyagi. O projeto se concentrou no processamento e distribuição de dados gerados por sistemas produtivos automatizados.

- O autor organizou, sob coordenação do Prof. Lucas Antonio Moscato, o projeto de um robô de inspeção de pequeno porte (denominado ROPET). O robô foi desenvolvido por alunos de iniciação científica; o trabalho tem como foco a geração de mão de obra capacitada em robótica. Participaram também desse projeto Fabio Tozeto Ramos e Marko Ackermann, alunos de mestrado, e Celso Emidio de Barros, funcionário da Universidade de São Paulo.

1.2.4 Breve descrição da pesquisa em dispositivos para automatização da marcha de pessoas com deficiências motoras

Técnicas de robótica voltadas ao auxílio de deficientes motores oferecem excelentes oportunidades de pesquisa; além disso, a criação de equipamentos nacionais pode ter impacto direto no bem-estar de uma parcela significativa da população brasileira.

Tendo em vista esses fatos, um projeto de pesquisa foi concebido em 1999 em parceria com a AACD (Associação de Assistência à Criança Defeituosa), uma instituição líder no atendimento a deficientes motores. O projeto consiste na implementação de um dispositivo de controle automático de elevação da junta do joelho em órteses; o objetivo é fazer com que o deficiente possa se movimentar com uma órtese em movimentos suaves, próximos dos movimentos de uma pessoa sem deficiência.

O projeto recebeu suporte da FAPESP na forma de bolsa de mestrado para o aluno Marko Ackermann. Foi construído um protótipo mecânico e foram realizados testes em um paciente da AACD; resultados foram publicados em [1, 2].

Capítulo 2

A teoria de conjuntos credais

Este capítulo descreve os aspectos fundamentais da teoria de conjuntos de medida de probabilidade. Conceitos básicos estão sumarizados na Seção 2.2. Esses conceitos podem ser encontrados na literatura e foram retirados de artigos do autor no *Artificial Intelligence Journal* [34] e no *Journal of Computational and Graphical Statistics* [31].

As seções 2.3 e 2.4 descreve os principais algoritmos para inferências com conjuntos de probabilidades; o texto segue artigo do autor no *International Journal of Approximate Reasoning* [33], incluindo resultados derivados no período 1999-2000 e publicados naquele artigo. Resultados novos sobre algoritmos para condicionantes com probabilidade zero, discutidos na Seção 2.5, foram derivados pelo autor no período 1999-2001 e publicados na conferência FLAIRS2002 [42].

2.1 Conjuntos credais em inteligência artificial

Desde o início da pesquisa em inteligência artificial, pesquisadores têm procurado representar e manipular adequadamente situações que envolvem incerteza. A teoria de probabilidades foi inicialmente considerada muito complexa computacionalmente, e foram procuradas outras soluções como teoria de Dempster-Shafer ou lógica difusa [103].

A teoria de probabilidades passou a ser amplamente utilizada em inteligência artificial durante a década de 80, quando foi criado o modelo de redes Bayesianas [90].

A teoria de probabilidades gradualmente se consolidou em inteligência artificial e hoje influencia a maioria das técnicas de tomada de decisão e de aprendizado de máquina [97]. Em inteligência artificial, o termo *teoria Bayesiana* é frequentemente utilizado como sinônimo de *teoria de probabilidades*; de forma geral, o adjetivo *Bayesiano* indica simplesmente que são utilizadas medidas de probabilidade. Embora o termo Bayesiano tenha conotações mais matizadas em outras áreas, esta dissertação utilizará este termo conforme o uso corrente em inteligência artificial.

Existem várias situações onde a utilização de uma única distribuição de probabilidades não é ideal: (1) análise de tomada de decisões por um grupo com opiniões divergentes; (2) análise do efeito de perturbações em valores de probabilidade, como feito no campo de *estatística robusta*; (3) análise de tomada de decisões quando o conhecimento de valores de probabilidade é impreciso ou vago, devido a imperfeições ou limitações dos tomadores de decisão [9, 85, 102, 113]. Nessas situações pode-se usar *conjuntos de probabilidades*. Por exemplo, pode-se dizer que "a probabilidade do evento A está no intervalo [0.2, 0.3]" — esse intervalo caracteriza um conjunto de probabilidades.

Várias teorias de inferência e decisão advogam o uso de conjuntos de probabilidade. Por exemplo, a *teoria quasi-Bayesiana* de Giron e Rios [63], a teoria de medidas internas/externas [108, 64, 96, 69], a teoria de probabilidades inferiores [105, 59, 11, 16], teorias de intervalos de probabilidade [67], a *teoria de conjuntos credais* [85], a *teoria de probabilidades imprecisas* baseada em axiomas de coerência [112], e mesmo teoria que combinam conjuntos de probabilidade com conjuntos de utilidade [100]. A teoria de conjuntos de probabilidade, utilizada neste trabalho, é equivalente em linhas gerais às teorias de conjuntos credais, de probabilidades imprecisas, e quasi-Bayesiana. O primeiro nome será adotado nesta dissertação.

A teoria de conjuntos de medidas de probabilidade tem enorme potencial por oferecer uma generalização axiomática e bem fundamentada à teoria de probabilidade. O objetivo da teoria de conjuntos de medidas de probabilidade, ao menos na forma apresentada neste trabalho, não é substituir a teoria de probabilidades, mas sim resolver problemas que não possuem solução satisfatória dentro da teoria de probabilidades.

2.2 Conceitos básicos

A teoria de conjuntos de medidas de probabilidade apareceu inicialmente nos escritos de H. Kyburg Jr. [77], C. Smith [105] e I. Levi [84]. Após uma evolução lenta e constante na década de 70, a década de 80 testemunhou enorme avanço devido ao crescimento da teoria de robustez Bayesiana [78, 9]. Em 1991, P. Walley publicou uma extensa obra contendo todos os aspectos teóricos conhecidos da teoria.

Um conjunto fechado de medidas de probabilidade é denominado *conjunto credal* [85]; a existência de tais conjuntos pode ser derivada axiomaticamente a partir de postulados sobre preferências [63]. O termo "'conjunto credal" será usado tanto para conjuntos de medidas de probabilidade quanto para conjuntos de distribuições de probabilidade e conjuntos de densidades de probabilidade.

Um conjunto credal definido por um conjunto de densidades $p(X)$ sobre a variável aleatória X é denotado por $K(X)$. Um vetor de variáveis $\mathbf{X}$ leva a um conjunto credal denotado por $K(\mathbf{X})$. As medidas de probabilidade que compõe um conjunto credal devem satisfazer algumas propriedades, ditas *axiomas de coerência* [18]. Uma medida P é coerente se satisfaz as seguintes condições:

1. $P(A) \geq 0$ para todo evento A;
2. a medida do evento universal é igual a 1;
3. para conjuntos disjuntos A e B, $P(A \cup B) = P(A) + P(B)$;
4. $P(A \cap B) = P(A|B) \times P(B)$.

Nesta dissertação o interesse está focado em conjuntos de probabilidade com um número finito de vértices; tais conjuntos são ditos *finitamente gerados* e são politopos no espaço de medidas de probabilidade [65]. Considere os seguintes exemplos, extraídos de [34]:

Exemplo 1 Um conjunto credal *ϵ-contaminado* é definido pelas densidades $p(X) = (1-\epsilon)r(X) + \epsilon q(X)$, onde $r(X)$ é uma densidade de probabilidade, $\epsilon \in (0,1)$ é um número real, e $q(X)$ indica qualquer densidade de probabilidade [9]. Alternativamente, um conjunto ϵ-contaminado contém as densidades $p(X) \geq (1-\epsilon)r(X)$ [11, 71]. Quando

X tem um número finito de valores, o conjunto ϵ-contaminado tem um número finito de vértices.

Exemplo 2 Um conjunto credal *de densidades limitadas* é definido pelas medidas $P(\cdot)$ tal que $L(A) \leq P(A) \leq U(A)$ para eventos A, sendo $L(A)$ e $U(A)$ medidas não-negativas, $L(\cup A) \leq 1$ e $U(\cup A) \geq 1$ [81, 83].

Dado um conjunto credal $K(X)$ e uma função limitada $f(X)$, *valores esperados inferiores e superiores* são respectivamente definidos por:

$$\underline{E}[f(X)] = \min_{P \in K} E_P[f(X)], \qquad \overline{E}[f(X)] = \max_{P \in K} E_P[f(X)],$$

onde $E_P[f(X)]$ é o valor esperado de $f(X)$ com relação à probabilidade P. Mínimos e máximos de valores esperados ocorrem apenas nos vértices de um conjunto credal [112]. Valores esperados inferiores e superiores se relacionam por $\underline{E}[f(X)] = -\overline{E}[-f(X)]$.

Da mesma forma, as *probabilidades inferior e superior* de um evento A são respectivamente definidas por:

$$\underline{P}(A) = \min_{P \in K} P(A), \qquad \overline{P}(A) = \max_{P \in K} P(A).$$

Probabilidades inferiores e superiores se relacionam por $\underline{P}(A) = 1 - \overline{P}(A^c)$. Deve-se notar que um cojunto credal sempre induz probabilidades inferiores e superiores únicas, mas um conjunto de probabilidades inferiores e superiores não define um único conjunto credal [112, Section 2.7].

O conjunto credal *condicional* $K(X|A)$ contém densidades condicionais $p(X|A)$. Esse conjunto, denominado *conjunto credal condicional*, é obtido em geral pela aplicação de regra de Bayes a cada elemento de um conjunto credal $K(X)$ [63, 85, 112]. Dado um conjunto credal $K(X)$ e uma função limitada $f(X)$, o valor esperado inferior de $f(X)$ condicional em um evento A é denotado por $\underline{E}[f(X)|A]$; se A tem probabilidade inferior maior que zero,

$$\underline{E}[f(X)|A] = \min_P E_P[f(X)|A] = \min_P \frac{E_P[I_A(X)f(X)]}{E_P[I_A(X)]},$$

onde $I_A(X)$ é a função indicadora do evento A. Se a probabilidade inferior de A é maior que zero, $\underline{E}[f(X)|A]$ é única solução da seguinte equação em λ, chamada *regra de Bayes generalizada* [112, Section 6.4]:

$$\underline{E}[(f(X) - \lambda)I_A(X)] = 0. \tag{2.1}$$

O cálculo de $\underline{E}[f(X)|A]$ quando $\underline{P}(A) = 0$ é discutido na Seção 2.5.

Um conjunto credal pode ser entendido como uma coleção de restrições sobre probabilidades. Considere variáveis com número finito de valores; cada valor esperado inferior $\underline{E}[f_i(X)|A_i]$ é uma restrição linear no espaço de medidas de probabilidade: $E_P[f_i(X)I_{A_i}(X)] \geq \underline{E}[f_i(X)|A_i]\, E_P[I_{A_i}(X)]$. Quando temos uma coleção de restrições de probabilidade, pode haver mais de um conjunto credal que satisfaz todas as restrições; cada um desses conjuntos credais é denominado uma *extensão* das restrições. O termo *extensão natural* indica o *maior* conjunto credal para um determinado conjunto de restrições. Ou seja, a extensão natural contém todas as medidas de probabilidade coerentes que satisfazem um determinado conjunto de restrições.

Um conjunto credal $K(X|Y)$ é *separadamente especificado* quando as restrições de probabilidade que definem $K(X|y_1)$ não interferem com as restrições que definem $K(X|y_2)$, para $y_1 \neq y_2$. De forma geral, uma coleção de conjuntos credais é separadamente especificada quando medidas de probabilidade podem ser independentemente selecionadas dos vários conjuntos credais condicionais [112, Chapter 6].

Há dois diferentes esquemas para representar conjuntos credais condicionais:

1. Como uma coleção de conjuntos credais *separadamente especificados* $K(X|Y = y)$, onde cada conjunto credal (para cada valor de Y) é definido de forma independente dos demais. A coleção de conjuntos $K(X|y)$ é denotada por $K(X|Y)$.

2. Como a especificação direta de um conjunto de funções $p(X|Y)$, chamado aqui de um conjunto credal condicional *extensivo*, e denotado por $\boldsymbol{L}(X|Y)$.

2.3 Valores esperados condicionais inferiores

Há três métodos básicos para especificar um conjunto credal: primeiro, pela especificação de um conjunto de vértices (cada um dos vértices é uma medida de probabilidade); segundo, pela especificação direta de restrições lineares que devem ser satisfeitas pelo conjunto credal; terceiro, pela utilização de alguma vizinhança topológica no espaço de medidas de probabilidade, como por exemplo conjuntos ϵ-contaminados ou *capacidades de Choquet* [15, 69, 117, 111].

Se um conjunto credal $K(X)$ é especificado como uma lista de vértices e A é um evento tal que $P(A) > 0$ para todos os vértices, o valor esperado condicional $E_p[f(X)|A]$ pode ser calculado para cada vértice de $K(X)$ através da regra de Bayes, e $\underline{E}[f(X)|A]$ será o máximo desses valores.

Se um conjunto credal é especificado diretamente por restrições lineares, o cálculo do valor esperado *não condicional* $\underline{E}[f(X)]$ é obtido pela minimização da função linear $E_P[f(X)]$ sujeita a restrições lineares — ou seja, é um problema de programação linear. Esta formulação remonta ao trabalho de Boole, conforme indicado e generalizado por Hailperin [68], e redescoberto em [12] e [88], entre outros [70]. Em inteligência artificial, esse tipo de formulação usualmente recebe o nome de *lógica probabilística* [88], da qual existem diversas variantes [8, 57, 110, 60, 86, 56]. O cálculo de um valor esperado condicional inferior $\underline{E}[f(X)|A]$ a partir de restrições lineares é significativamente mais complexo que o cálculo de $\underline{E}[f(X)]$, e será denominado *Problema A*.

Se um conjunto credal é especificado através de alguma vizinhança topológica ou capacidade de Choquet, o cálculo de valores esperados inferiores $\underline{E}[f(X)]$ é em geral feito por fórmulas fechadas, enquanto que o cálculo de valores esperados condicionais inferiores $\underline{E}[f(X)|A]$ requer algoritmos mais sofisticados. Essa última situação será denominada *Problema B*.

O restante dessa seção aborda a solução dos Problemas A e B. Assume-se que o valor esperado inferior $\underline{E}[f(X)|A]$ é calculado para um evento A com $\underline{P}(A) > 0$; algoritmos para a situação $\underline{P}(A) = 0$ são discutidos na Seção 2.5.

O *algoritmo de Lavine* é um esquema iterativo para cálculo de valores esperados condicionais [81]. Defina $\underline{\mu}_0 = \inf f(X)I_A(X)$ and $\overline{\mu}_0 = \sup f(X)I_A(X)$. Defina também $M(\mu) = \underline{E}[(f(X) - \mu)I_A(X)]$; note que $M(\mu)$ tem que ser igual a zero para algum μ no intervalo $[\underline{\mu}_0, \overline{\mu}_0]$. Reduza esse intervalo através da seguinte procedimento (para $i \geq 0$):

1. Pare se $|\overline{\mu}_i - \underline{\mu}_i| < \epsilon$ para algum valor positivo ϵ;
2. Escolha μ_i in $[\underline{\mu}_i, \overline{\mu}_i]$, e se $M(\mu_i) > 0$, tome $\underline{\mu}_{i+1} = \mu_i$ e $\overline{\mu}_{i+1} = \overline{\mu}_i$; se $M(\mu_i) < 0$, tome $\underline{\mu}_{i+1} = \underline{\mu}_i$ e $\overline{\mu}_{i+1} = \mu_i$.

O teorema seguinte demonstra que o valor $M(\mu_i)$ pode ser usado para determinar

quando parar o procedimento.

Teorema 1 *Para um evento A tal que $\underline{P}(A) > 0$, se $|M(\mu)| \leq \epsilon \underline{P}(A)$, então*

$$|\mu - \underline{E}[f(X)|A]| \leq \epsilon.$$

A prova desse teorema pode ser encontrada em artigo publicado pelo autor no *International Journal of Approximate Reasoning* [33].

O algoritmo de Lavine é imediato para o Problema A, quando variáveis são categóricas (isto é, tem um número finito de valores) e para Problema B. No primeiro caso, o algoritmo se reduz a uma sequência de programas lineares (um para cada valor de μ_i) [82].

Um outro esquema iterativo baseado na regra de Bayes generalizada foi proposto por Walley [112, Nota 6.4.1]. Nesse esquema, $\underline{E}[f(X)|A]$ é obtida iterando $\mu_{i+1} = \mu_i + 2\underline{E}[(f(X) - \mu_i)I_A(X)] / (\overline{E}[I_A(X)] + \underline{E}[I_A(X)])$. Este algoritmo pode ser facilmente aplicado ao Problema B. Walley provou que seu algoritmo tem convergência linear: $\epsilon_{i+1} = \delta\epsilon_i$, onde $\delta = (\overline{P}(A) - \underline{P}(A))/(\overline{P}(A) + \underline{P}(A))$ e ϵ_i é o erro no passo i. Note que o algoritmo de Lavine também tem convergência linear (se bissecção for utilizada, o algoritmo tem erro $\epsilon_{i+1} = (1/2)\epsilon_i$). Para Problema B, o algoritmo de Walley é uma escolha melhor que o algoritmo de Lavine para $\delta < 1/2$; ou seja, quando $3\underline{P}(A) > \overline{P}(A)$.

De forma geral, o Problema B pode ser entendido como a busca numérica da solução da equação (2.1). Desse ponto de vista, é evidente que convergência linear não é a melhor solução possível; existem métodos com convergência super-linear, como o método da secante ou o método de Brent, que devem ser utilizados [92]. Tais métodos exigem em geral um valor inicial relativamente preciso. Assim sendo, podemos resumir a discussão do Problema B como segue:

Nota 1 *A melhor solução para Problema B é usar um método super-linear para resolver a regra de Bayes generalizada (por exemplo, método de Brent), usando o método de Lavine ou método de Walley (dependendo do valor de δ) para atingir uma vizinhança da solução.*

Retornando ao Problema A, a melhor solução existente não é seguir algoritmos iterativos, mas sim aplicar métodos de programação linear fracional [98]. Existem dois

algoritmos clássicos para solução de programas lineares fracionais, e só recentemente esses algoritmos têm sido associados a conjuntos de probabilidade [58, 76, 87, 89]. O primeiro método é denominado *algoritmo de Dinkelbach ou Jagannatham*, e é virtualmente idêntico ao algoritmo de Lavine. O segundo método é conhecido como *transformação de Charnes-Cooper*, e é o método mais indicado na prática.

A transformação de Charnes-Cooper transforma o cálculo de $\underline{E}[f(X)|A]$ em um programa linear:

$$\begin{aligned} &\min \textstyle\sum_X I_A(x) f(x) Q(x); \text{ sujeito a} \\ &\sum_X I_A(x) Q(x) = 1, \qquad Q(x) \geq 0, \\ &\sum_X I_{A_i}(x) \left(f_i(x) - \underline{E}[f_i(X)|A_i]\right) Q(x) \geq 0, \end{aligned} \tag{2.2}$$

para todas as funções $f_i(X)$, $I_{A_i}(X)$ usadas para especificar o conjunto credal.

Assim, temos:

Nota 2 *A melhor solução para o Problema A é a transformação de Charnes-Cooper expressa em (2.2).*

Além dos casos indicados, algumas vizinhanças topológicas apresentam expressões fechadas para cálculo de probabilidades inferiores condicionais [15, 55, 111].

2.4 Verossimilhança e valores esperados condicionais inferiores

Problemas que envolvem inferências sobre uma variável Y a partir de observações de uma variável X em geral manipulam dois conjuntos credais. Primeiro, um conjunto credal $K(Y)$ que indica crenças a priori (antes das observações). Segundo, um conjunto credal $K(X|Y)$, denominado *verossimilhança*, que modela as observações. Como exemplo, a área de estatística robusta usualmente utiliza um conjunto credal $K(Y)$ e uma única distribuição $p(X|Y)$ para modelar situações onde crenças iniciais não são muito sólidas [9, 78, 10, 116].

Considere agora o Problema A para variáveis discretas, com a presença de uma única função de verossimilhanca $L_x(Y) = p(x|Y)$. O algoritmo de Lavine pode ser aplicado diretamente; o cálculo de $\underline{E}[f(Y)|x]$ usa $M(\mu) = \underline{E}[(f(Y) - \mu)L_x(Y)]$ [112]. Outro algoritmo proposto para esta mesma situação é o algoritmo de White-Snow [106, 119]. Para descrever o algoritmo, considere que o conjunto credal $K(Y)$ é especificado por restrições lineares

$$\mathbf{A}[P(Y = y_1) \dots P(Y = y_n)]^T \leq \mathbf{B},$$

onde $\mathbf{A}$ é uma matriz e $\mathbf{B}$ é um vetor de dimensões apropriadas. Defina os vetores α por $\alpha_i = P(Y = y_i)$, β por $\beta_i = P(X = x|y_i)$, e f por $f_i = f(y_i)$, e a matriz $\mathbf{C} = \mathbf{A} - \mathbf{B1}$ (onde $\mathbf{1}$ é uma linha de 1's). Então:

$$\underline{E}[f(Y)|x] = \min_{\alpha} \left[\frac{\sum_i f_i \alpha_i \beta_i}{\sum_j \alpha_j \beta_j} \right], \quad \text{sujeito a} \quad \mathbf{C}\alpha \leq 0, \sum_i \alpha_i = 1, \alpha_i \geq 0. \tag{2.3}$$

O algoritmo de White-Snow pode ser entendido como uma transformação que "lineariza" um problema fracional, sendo bastante semelhante à transformação de Charnes-Cooper. O algoritmo de White-Snow poderia ser aplicado mesmo para solução do Problema A original (sem verossimilhança); de qualquer forma, a tranformação de Charnes-Cooper é numericamente mais estável por não envolver divisões por valores de probabilidade.

Considere agora a situação em que mais de uma função de verossimilhança pertence a $K(X|Y)$. Ou seja, que tenhamos um conjunto credal $K(Y)$ e um conjunto credal $K(X|Y)$, ambos contendo várias funções. Essa possibilidade, de importância prática, têm sido raramente considerada na literatura [81, 91, 112].

O próximo teorema prova que algoritmos podem restringir atenção a máximos e mínimos da função de verossimilhança quando manipulando conjuntos de de funções de verossimilhança. O teorema usa os conceitos de *verossimilhanças inferiores e superiores*, ou seja, para conjuntos credais $K(X|Y)$, a *verossimilhanças inferior* $L_x(Y)$ é uma função definida como $L_x(y) = \underline{P}(X = x|y) = \min_{p(X|y) \in K(X|y)} P(X = x|y)$, e a *verossimilhanças superior* $U_x(Y)$ é uma função definida como $U_x(y) = \overline{P}(X = x|y) = \max_{p(X|y) \in K(X|y)} P(X = x|y)$.

Teorema 2 (Walley [112, Seção 8.5.3]) *Considere uma função limitada $f(Y)$ e suponha que $K(X|Y)$ e $K(X)$ são separadamente especificados. Se $\underline{P}(X = x) > 0$, então*

$\underline{E}[f(Y)|x]$ é o único valor de μ tal que $\underline{E}[(f(Y)-\mu)\,p_\mu(x|Y)]=0$, onde $p_\mu(x|y)$ é igual a $U_x(y)$ se $f(y)\geq\mu$ e igual a $L_x(y)$ se $f(y)<\mu$.

O teorema demonstra que, se $\underline{P}(X=x)>0$,

$$\underline{E}[f(Y)|x]=\min\left(E_p[f(Y)p_\mu(x|Y)]\,/E_p[p_\mu(x|Y)]\right),$$

onde a minimização é com respeito a
(i) $\mu\in[\inf f(Y)I_x(X),\sup f(Y)I_x(X)]$, e
(ii) $p(Y)\in K(Y)$.
Portanto, é possível usar um esquema iterativo como o algoritmo de Lavine, utilizando uma "verossimilhança" $p_\mu(x|Y)$ que varia a cada iteração do algoritmo. Cada passo do algoritmo envolve o cálculo de $M(\mu)=\underline{E}[(f(Y)-\mu)p_\mu(x|Y)]$.

Um método mais satisfatório para o cálculo de valores esperados inferiores $\underline{E}[f(Y)|x]$, a partir de conjuntos credais finitamente gerados e separadamente especificados $K(Y)$ e $K(X|Y)$, é o seguinte. Primeiro defina dois vetores, α' and α'', cada um com o mesmo comprimento de α. Defina então o seguinte programa linear fracional:

$$\underline{E}[f(Y)|x]=\min_{\alpha',\alpha''}\left[\frac{\sum_i\left(f_iL_x(y_i)\alpha'_i+f_iU_x(y_i)\alpha''_i\right)}{\sum_j\left(L_x(y_j)\alpha'_j+U_x(y_j)\alpha''_j\right)}\right],\tag{2.4}$$

sujeito a: $\mathbf{C}\ (\alpha'+\alpha'')\leq 0$, $\sum_i(\alpha'_i+\alpha''_i)=1$, $\alpha'_i\geq 0$, $\alpha''_i\geq 0$. Para cada i, os valores de α' e α'' que minimizam a função objetivo satisfazem $\alpha'_i=0$ ou $\alpha''_i=0$ para cada i, automaticamente selecionando o valor de verossimilhança inferior ou superior. A partir desse programa fracional, a transformação de Charnes-Cooper pode ser usada.

2.5 Condicionando com probabilidade inferior nula

As seções anteriores discutiram algoritmos para cálculo de $\underline{E}[f(X)|A]$ quando $\underline{P}(A)>0$. Nos próximos capítulos dessa dissertação, todos os resultados assumem que eventos condicionantes tem probabilidade inferior maior que zero. Nesta seção discutimos a possibilidade do evento condicionante ter probabilidade inferior nula, mostrando como essa situação pode ser enfrentada. Devido à maior complexidade do tema, vamos considerar apenas o caso de variáveis categóricas e de restrições lineares aqui.

Suponha então que um evento B tem probabilidade inferior igual a zero e probabilidade superior maior que zero. Ou seja, é "aceitável" que a probabilidade de B seja zero. É preciso obter um valor para $\underline{E}[f(X)|B]$ que seja coerente com outros valores esperados inferiores $\{\underline{E}[f_i(X)|B_i]\}_{i=1}^{k}$ — esses valores justamente definem o conjunto credal $K(X)$. Uma solução frequentemente usada é focar apenas nas medidas de probabilidade que colocam probabilidade maior que zero em B:

$$\underline{E}[f(X)|B] = \min_{P:P(B)>0} E_P[f(X)|B]\,. \tag{2.5}$$

A idéia é que devemos descartar modelos que coloquem probabilidade zero sobre B a partir do momento que B seja observado [118]. No caso de $\overline{P}(B) = 0$ (quando todas as medidas de probabilidade no conjunto credal levam a $P(B) = 0$), pode-se então declarar que $\underline{E}[f(X)|B]$ é indefinido. Note que a função $Q(X)$ em (2.2) é proporcional a uma medida de probabilidade que coloca probabilidade maior que zero em B. Consequentemente, o programa linear (2.2) automaticamente descarta modelos onde $P(B) = 0$. Quando $\overline{P}(B) = 0$, o programa linear não tem solução factível; nesses casos deve-se indicar que o modelo deve ser revisado.

Um ponto de vista diferente, que não descarta modelos, considera a "extensão natural" como ponto de partida (Seção 2.2) — ou seja, parte do princípio que devemos construir o maior conjunto de medidas coerentes de probabilidades [54]. A definição de medida coerente pode ser encontrada na Seção 2.2. Em uma medida coerente, não há necessidade de termos igualdade entre $P(A|B)$ e $P(A \cap B)/P(B)$; há apenas a restrição que $P(A \cap B)$ seja igual a $P(A|B) \times P(B)$. Se $P(B)$ é igual a zero, então $P(A \cap B)$ e $P(A|B) \times P(B)$ são *sempre* iguais, independentemente de $P(A|B)$. Devido a isso, o cálculo de valores esperados inferiores para extensões naturais é um problema não-linear quando eventos condicionantes tem probabilidade inferior nula. A não-linearidade de inferências na presença de eventos com probabilidade inferior nula introduz várias dificuldades. Walley, Pelessoni and Vicig derivam algoritmos para resolver este problema [114], mas não chegam a soluções inteiramente satisfatórias. Em primeiro lugar, os métodos de Walley, Pelessoni e Vicig não detectam quando a extensão natural pode ser obtida diretamente pela Expressão (2.2); deveria ser possível detectar automaticamente quando eventos condicionantes tem probabilidade inferior maior que zero. Além disso, os métodos requerem sequências de programas lineares, incorrendo custos para a montagem dos programas lineares.

Uma solução completa para essas dificuldades pode ser construída através da in-

vestigação de extensões naturais. O autor produziu um algoritmo que calcula valores esperados inferiores que tem a seguintes propriedades: quando o evento condicionante tem probabilidade inferior maior que zero, o algoritmo leva a um único programa linear; caso contrário, o algoritmo segue uma sequência de operações de pivotamento que são equivalentes à sequência de programas lineares de WPV, mas com custo computacional menor [42]. No momento, este algoritmo oferece o método mais geral e eficiente para cálculo de valores esperados inferiores.

2.6 Discussão

Este capítulo apresentou a teoria de conjuntos credais, enfatizando aspectos algorítmicos. Foram discutidos algoritmos para cálculo de valores esperados inferiores em uma série de situações. Alguns desses algoritmos foram derivados pelo autor no período 1998-2002; em particular, o algoritmo resumido pela Expressão (2.4) é o método mais eficiente para cálculo de valores esperados condicionais inferiores na presença de verossimilhanças "imprecisas", desde que condicionantes tenham probabilidade inferior maior que zero. O algoritmo para manipulação de eventos condicionantes com probabilidade inferior igual a zero, discutido na Seção 2.5, também é o mais eficiente em sua categoria.

Capítulo 3

Independência para conjuntos credais

Este capítulo apresenta os principais conceitos de independência e independência condicional para conjuntos credais. Os conceitos apresentados são: irrelevância epistêmica, independência epistêmica, independência forte, e condição de Kuznetsov. A Seção 3.2 demonstra que esses conceitos de independência não levam necessariamente à convergência de opiniões em experimentos repetitivos; para obter convergência de opiniões, o conceito apropriado é o de *permutabilidade*. Uma vez determinado o escopo do conceito de permutabilidade, o foco retorna aos conceitos de independência. As Seções 3.3, 3.4 e 3.5 analisam conceitos de independência com base nas *propriedades de semi-grafóides* — um conjunto de propriedades que são usualmente associadas ao conceito abstrato de "independência". A conclusão é que independência forte e independência epistêmica são satisfatórias do ponto de vista de grafóides, mas a condição de Kuznetsov enfrenta dificuldades.

O material aqui apresentado está contido em artigos publicados no *International Journal of Approximate Reasoning* [33], na *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* 1999 [32], na *Conference on Uncertainty in Artificial Intelligence* 2000 [35], e no *International Symposium on Imprecise Probabilities and Their Applications* 2001 [48, 41]. Alguns resultados são novos e estão contidos em artigos recentemente submetidos para avaliação.

3.1 Conceitos de independência

Uma das dificuldades correntes na teoria de conjuntos de probabilidade é a falta de consenso sobre qual é a definição matemática correspondente ao conceito de independência [20, 52]. Note que, na teoria de probabilidades usual, a definição de independência é relativamente simples: variáveis X e Y são *estocasticamente independentes* se $p(X|y) = p(X)$ para todo valor y; ou, alternativamente, $p(X, Y) = p(X)\,p(Y)$.

Vários autores tem procurado adaptar o conceito de independência usual a conjuntos de probabilidade. Os conceitos mais importantes são apresentados a seguir. Neste capítulo, como no restante deste trabalho, assume-se que todos os eventos tem probabilidade inferior maior que zero, evitando assim casos especiais nas definições e resultados.

3.1.1 Independência epistêmica

O significado intuitivo de "Y é irrelevante a X" é que a observação de Y não deve mudar a incerteza sobre X. Sendo a incerteza sobre X expressa por um conjunto credal, podemos argumentar que a irrelevância de Y sobre X deve significar que $K(X|y)$ é equivalente a $K(X)$ para todo valor de Y [21, 28].

Definição 1 *Variável Y é* epistemicamente irrelevante *a X condicional em Z se $K(X|z)$ e $K(X|y, z)$ tem o mesmo fecho convexo para todo valor de Y e Z.*

Note que dois conjuntos credais são equivalentes, em termos de valores esperados, quando seus fechos convexos são iguais.

Alternativamente, podemos definir irrelevância epistêmica usando valores esperados inferiores: Y é *epistemicamente irrelevante* a X quando $\underline{E}[f(X)|Y = y] = \underline{E}[f(X)]$ para todos os valores de Y e todas as funções limitadas de X (esse conceito pode ser formulado de maneira condicional, deixando-se uma variável Z como condicionante). Esta é a formulação original do conceito de irrelevância epistêmica por Walley [112, Section 9.3].

Ao contrário de irrelevância, o conceito de independência representa uma relação simétrica. Por isso, temos:

Definição 2 *Variáveis X e Y são* epistemicamente independentes *condicional em Z se X é epistemicamente irrelevante a Y condicional em Z e se Y é epistemicamente irrelevante a X condicional em Z.*

3.1.2 Independência forte

Um conceito de independência bastante utilizado para conjuntos credais é o seguinte:[1]

Definição 3 *Variáveis X e Y são* fortemente independentes *condicional em Z quando todo vértice de $K(X,Y)$ satisfaz independência estocástica entre X e Y condicional em Z.*

Independência forte é razoável quando um conjunto credal é interpretado como um conjunto contendo uma probabilidade "correta", e não como uma representação exaustiva de incertezas [112]. Independência forte é também razoável quando uma coleção de especialistas, responsáveis pela construção de um conjunto credal, concorda que todas as probabilidades de um modelo devem forçosamente satisfazer independência estocástica.

A despeito do apelo intuitivo do conceito de independência forte, não há uma justificativa direta na literatura para este conceito — ou seja, uma justificativa que não utilize independência estocástica como ponto de partida. Definições de independência forte usualmente assumem independência estocástica; isto é pouco desejável pois a teoria de conjuntos credais pretende ser um modelo mais básico de incerteza que a teoria de probabilidades tradicional. Independência forte se apresenta como uma generalização puramente matemática das propriedades de independência estocástica. Essa situação será abordada no Capítulo 4, onde uma solução será proposta.

[1] Este conceito é chamado de *independência na seleção* por Couso et al [20], que empregam o termo "independência forte" para o conjunto credal que satisfaz Definição 3. Aqui o termo "independência forte" se refere a um conceito de independência, e o maior conjunto credal satisfazendo a definição é denominado *extensão forte* (Seção 4.2).

3.1.3 Condição de Kuznetsov

A condição de Kuznetsov procura caracterizar a situação em que duas variáveis são consideradas "independentes". A idéia é considerar funções do tipo $f(X)g(Y)$; caso os valores esperados se "decomponham de forma apropriada", então as variáveis são ditas "independentes".[2] Para formalizar a idéia, considere uma função $h(Z)$ qualquer e um conjunto credal $K(Z)$. O conjunto credal determina um intervalo para os valores esperados de $h(Z)$: $\mathbf{E}[h(Z)] = [\underline{E}[h(Z)], \overline{E}[h(Z)]]$. Note que dois intervalos $[a, b]$ e $[c, d]$ podem ser multiplicados, levando ao intervalo $[\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)]$.

De acordo com a condição de Kuznetsov, variáveis X e Y são independentes se

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)] \times \mathbf{E}[g(Y)],$$

para todas as funções limitadas $f(X)$ e $g(Y)$, e usando multiplicação intervalar. Essa condição é equivalente a:

$$\underline{E}[f(X)g(Y)] = \min \begin{pmatrix} \underline{E}[f(X)]\,\underline{E}[g(Y)]\,, \underline{E}[f(X)]\,\overline{E}[g(Y)]\,, \\ \overline{E}[f(X)]\,\underline{E}[g(Y)]\,, \overline{E}[f(X)]\,\overline{E}[g(Y)] \end{pmatrix}. \tag{3.1}$$

A generalização natural da condição de Kuznetsov para valores esperados condicionais é a seguinte. Considere que duas variáveis X e Y são independentes condicional em um evento A se

$$\mathbf{E}[f(X)g(Y)|A] = \mathbf{E}[f(X)|A] \times \mathbf{E}[g(Y)|A],$$

para todas as funções limitadas $f(X)$ e $g(Y)$. Assim sendo, X e Y são independentes (de acordo com a condição de Kuznetsov) condicional em Z se

$$\mathbf{E}[f(X)g(Y)|z] = \mathbf{E}[f(X)|z] \times \mathbf{E}[g(Y)|z],$$

para todo valor z.

3.2 Independência e permutabilidade

Uma das aplicações básicas do conceito de independência estocástica tradicional ocorre no contexto de experimentos repetidos. Suponha que uma sequência de observações

[2] A condição foi proposta pelo pesquisador russo V. P. Kuznetsov. Infelizmente, Kuznetsov faleceu em um acidente no início de 1998. A maior parte do trabalho de Kuznetsov está contido no livro *Interval Statistical Models*, publicado em 1991 em russo [79].

$X_1, \ldots, X_n$ está sendo coletada, sendo que as observações são independentes e modeladas identicamente pela densidade $p(X_k|\theta)$ — a função de verossimilhança do experimento. Um resultado clássico em estatística é que, à medida que mais e mais observações são coletadas, a densidade $p(\Theta|X_1, \ldots, X_n)$ se concentra em torno do valor θ. Esse fenômeno é ocasionalmente referido como *convergência de opiniões Bayesianas*, e é uma consequência direta da suposição de independência entre observações [99].

Esta seção investiga o comportamento de sequências de observações com relação a conjuntos de probabilidade. Note primeiro que os conceitos de independência apresentados não restringem de nenhuma maneira a forma como densidades são combinadas de experimento a experimento; a partir disso obtemos o seguinte resultado, onde "independência" pode significar qualquer um dos conceitos discutidos anteriormente:

Teorema 3 *Para uma sequência de observações independentes, as verossimilhanças inferior e superior são dadas respectivamente por*

$$U_{X_1,\ldots,X_n}(\Theta) = \prod_{k=1}^{n} U_{X_k}(\Theta) \quad e \quad L_{X_1,\ldots,X_n}(\Theta) = \prod_{k=1}^{n} L_{X_k}(\Theta).$$

A prova deste teorema pode ser encontrada em artigo publicado pelo autor no *Journal of Computational and Graphical Statistics* [31].

O exemplo a seguir ilustra o tipo de fenômeno que pode ocorrer com conjuntos credais em experimentos repetidos, quando a única restrição é a de independência entre experimentos. Note que não há formação de "consenso".

Exemplo 3 Considere uma variável discreta Θ com N valores. Um grupo de especialistas estabelece um conjunto credal $K(\Theta)$ que representa as crenças sobre Θ antes dos experimentos. O conjunto $K(\Theta)$ é tal que $\underline{P}(\Theta = \theta_j) > 0$ para todo θ_j. As observaç oes tem um número finito de valores. Os especialistas concordam que todas as observações são independentes e satisfazem o mesmo modelo $K(X_k|\Theta)$. Além disso, os especialistas notam que $\overline{P}(X_k|\theta_j) > \underline{P}(X_k|\theta_j) > 0$ para todo θ_j, e $\overline{P}(X_k|\theta_i) > \underline{P}(X_k|\theta_j)$ para todo $j \neq i$. Um terceiro grupo de especialistas então coleta uma sequência de observações X_k. Para sua surpresa, os especialistas notam que $\overline{P}(\theta_i|X_1, \ldots, X_n)$ tende a um e $\underline{P}(\theta_i|X_1, \ldots, X_n)$ tende a zero (a prova desse fato está no artigo citado anteriormente [31]).

Muitos conjuntos podem ser facilmente construídos que satisfazem as características descritas no exemplo. No entanto, o desfecho é surpreendente pois os especialistas perdem qualquer nível de consenso existente em vista das observações! Outros exemplos com propriedades similares podem ser encontrados na literatura [101, 115], mas o Exemplo 3 produz uma situação extrema em que todas as crenças se tornam "vácuas" a partir de observações.

O fenômeno ilustrado pelo Exemplo 3 é causado pela ausência de restrições "entre experimentos". Isso ocorre pois os conceitos de independência existentes para conjuntos credais são voltados para relações internas a um modelo. Para modelar experimentos repetitivos, precisamos esclarecer como os experimentos se relacionam; para tanto, uma propriedade importante é a *permutabilidade*.[3]

O conceito de permutabilidade é o seguinte. Considere um vetor de observações $\mathbf{X} = [X_1, \ldots, X_N]$, e indique por $\mathcal{P}(\mathbf{X})$ um vetor cujos elementos são uma permutação do vetor $\mathbf{X}$. Permutabilidade ocorre se $\underline{E}[\mathcal{P}(\mathbf{X}) - \mathbf{X}] = 0$. Ou seja, a diferença entre um vetor $\mathbf{X}$ e uma permutação desse vetor tem valor esperado nulo (não há o que "ganhar" trocando a ordem das variáveis).

Permutabilidade de $\mathbf{X}$ implica que:

$$\overline{E}[\mathcal{P}(\mathbf{X}) - \mathbf{X}] = -\underline{E}[-(\mathcal{P}(\mathbf{X}) - \mathbf{X})] = -\underline{E}[\mathbf{X} - \mathcal{P}(\mathbf{X})] = -\underline{E}\left[\mathcal{P}^{-1}(\mathbf{Y}) - \mathbf{Y}\right],$$

onde $\mathbf{Y}$ é $\mathcal{P}(\mathbf{X})$ e $\mathcal{P}^{-1}$ indica a permutação inversa de $\mathcal{P}$. Note que uma permutação inversa sempre é uma permutação, e portanto $\overline{E}[\mathcal{P}(\mathbf{X}) - \mathbf{X}] = 0$. Isso implica que todas as medidas de probabilidade no conjunto credal em questão satisfazem $P(\mathcal{P}(\mathbf{X})) = P(\mathbf{X})$. Isso caracteriza a permutabilidade de cada medida no conjunto credal.

Permutabilidade requer apenas que uma sequência de observações seja tal que qualquer mudança na ordem de observações não afete as conclusões a respeito da sequência. Mais importante, o fato de cada medida no conjunto credal satisfazer permutabilidade leva ao resultado de convergência desejado: uma sequência permutável de observações levará a uma distribuição limite, que sumariza o "consenso" entre todas as distribuições do conjunto credal [112]. Ou seja, a imprecisão traduzida pelo "volume" de um conjunto credal tende a zero quando o número de observações cresce indefinidamente.

[3] O conceito aqui discutido é denominado *exchangeability* por Walley [112]; infelizmente Walley também propõe um conceito diferente, menos relevante na prática, denominado *permutability*. As duas palavras da língua inglesa tem a mesma tradução na língua portuguesa.

Portanto, o método mais adequado para modelagem de sequências de observações, tanto para conjuntos credais quanto para medidas de probabilidade, é a suposição de permutabilidade entre observações. Essa suposição é simples, intuitiva, e tem excelentes propriedades. Por isso, o restante desse trabalho não se ocupará de sequências de observações, procurando focar em relações de independência "intra-modelo".

3.3 Independência forte e semi-grafóides

O objetivo desta seção e das seções seguintes é estudar o comportamento dos conceitos de independência apresentados, tendo como base as chamadas "propriedades de semi-grafóides". Essas propriedades foram propostas na literatura de Estatística e Inteligência Artificial como modelos abstratos do conceito de irrelevância e de independência [51, 90]. De fato, muitos conceitos importantes de independência, como separação de nós em grafos e independência probabilística, são semi-grafóides.

Uma relação ternária $X \perp\!\!\!\perp Y \,|\, Z$ é denominada um *semi-grafóide* quando satisfaz as seguintes propriedades [62, 90, 51, 107]:

Simetria: $X \perp\!\!\!\perp Y \,|\, Z \Rightarrow Y \perp\!\!\!\perp X \,|\, Z$

Redundância: $X \perp\!\!\!\perp Y \,|\, X$

Decomposição: $X \perp\!\!\!\perp (W, Y) \,|\, Z \Rightarrow X \perp\!\!\!\perp Y \,|\, Z$

União fraca: $X \perp\!\!\!\perp (W, Y) \,|\, Z \Rightarrow X \perp\!\!\!\perp W \,|\, (Y, Z)$

Contração: $X \perp\!\!\!\perp Y \,|\, Z \;\&\; X \perp\!\!\!\perp W \,|\, (Y, Z) \Rightarrow X \perp\!\!\!\perp (W, Y) \,|\, Z$

Dawid sugere que a notação $X \perp\!\!\!\perp Y \,|\, Z$ deve ser interpretada como indicando que "qualquer informação sobre Y é irrelevante para a incerteza sobre X" [51]. No contexto de conjuntos credais, deve-se tomar grafóides como relações entre variáveis W, X, Y e Z.

Em relação ao conceito de independência forte, temos:

Teorema 4 *Independência forte é um semi-grafóide.*

Prova. Simetria e redundância são imediatas pois todo vértice de $K(X,Y)$ satisfaz essas propriedades. Decomposição segue do fato que a marginalização de um conjunto credal é dada pela marginalização de cada vértice do conjunto; portanto se os vértices satisfazem $p(W,X,Y|Z) = p(X|Z)\,p(W,Y|Z)$, o marginal $p(X,Y|Z)$ satisfaz $\sum_W p(X|Z)\,p(W,Y|Z) = p(X|Z)\,p(Y|Z)$. União fraca segue do mesmo raciocínio: se $p(W,X,Y|Z)$ é igual a $p(X|Z)\,p(W,Y|Z)$, então o condicional $p(X,Y|W,Z)$ satisfaz $p(X|Z)\,p(W,Y|Z)\,/p(Y|Z) = p(X|Z)\,p(W|Y,Z)$. Contração segue do fato que, dadas as suposições de contração, todo vértice de $K(W,X,Y|Z)$ satisfaz $p(X|W,Y,Z) = p(X|Y,Z) = p(X|Z)$, e portanto todo vértice satisfaz independência estocástica de X e (W,Y). QED

3.4 Independência epistêmica e semi-grafóides

Como o conceito de irrelevância epistêmica fundamenta o conceito de independência epistêmica, e como o conceito de irrelevância epistêmica não é simétrico, é necessário distinguir entre diferentes versões das propriedades de semi-grafóides. Assim, a propriedade de redundância tem duas versões: $X \perp\!\!\!\perp Y \,|\, X$ e $Y \perp\!\!\!\perp X \,|\, X$. Propriedades de decomposição e união fraca tem quatro versões, enquanto contração tem oito possíveis versões, obtidas trocando as posições das variáveis. Quando as variáveis estão na posição indicada na seção anterior, a propriedade recebe o adjetivo "direta"; se as variáveis estão trocadas em todas as relações, a propriedade recebe o adjetivo "reversa". Por exemplo, temos:

Contração direta: $X \perp\!\!\!\perp Y \,|\, Z \;\&\; X \perp\!\!\!\perp W \,|\, (Y,Z) \Rightarrow X \perp\!\!\!\perp (W,Y) \,|\, Z$

Contração reversa: $Y \perp\!\!\!\perp X \,|\, Z \;\&\; W \perp\!\!\!\perp X \,|\, (Y,Z) \Rightarrow (W,Y) \perp\!\!\!\perp X \,|\, Z$.

Há outras seis formas da propriedade de contração, as quais não recebem nomes específicos.[4]

O próximo teorema resume as propriedades de semi-grafóides que são satisfeitas pelo conceito de irrelevância epistêmica. Note que aqui, como no restante da dissertação,

[4]Uma situação similar é encontrada no trabalho de Galles e Pearl [61], em relação a axiomas de causalidade. Em virtude da falta de simetria em relações causais, Galles e Pearl introduzem versões das propriedades de grafóides denominadas "da direita" e "da esquerda".

assume-se que todos os eventos tem probabilidade inferior maior que zero.

Teorema 5 *Tomando $X \perp\!\!\!\perp Y \,|\, Z$ como "X é irrelevante a Y condicional em Z, então irrelevância epistêmica satisfaz as seguintes propriedades:*

Redundância direta: $X \perp\!\!\!\perp Y \,|\, X$

Redundância reversa: $Y \perp\!\!\!\perp X \,|\, X$

Decomposição direta: $X \perp\!\!\!\perp (W, Y) \,|\, Z \Rightarrow X \perp\!\!\!\perp Y \,|\, Z$

Decomposição reversa: $(W, Y) \perp\!\!\!\perp X \,|\, Z \Rightarrow Y \perp\!\!\!\perp X \,|\, Z$

União fraca direta: $X \perp\!\!\!\perp (W, Y) \,|\, Z \Rightarrow X \perp\!\!\!\perp W \,|\, (Y, Z)$

União fraca reversa: $(W, Y) \perp\!\!\!\perp X \,|\, Z \Rightarrow W \perp\!\!\!\perp X \,|\, (Y, Z)$.

Contração reversa: $Y \perp\!\!\!\perp X \,|\, Z \& W \perp\!\!\!\perp X \,|\, (Y, Z) \Rightarrow (W, Y) \perp\!\!\!\perp X \,|\, Z$.

A prova deste teorema está em artigo publicado no *Second International Symposium on Imprecise Probabilities and Their Applications* [48]. Naquele artigo também se encontram exemplos que mostram como todas as outras versões das propriedades semi-grafóides podem ser violadas por irrelevância epistêmica.

Como independência epistêmica é uma versão "simetrizada" de irrelevância epistêmica, as propriedades do primeiro conceito seguem das propriedades do segundo:

Teorema 6 *Independência epistêmica satisfaz simetria, redundância, decomposição e união fraca.*

Embora exemplos de violação da propriedade de contração estejam disponíveis no artigo mencionado anteriormente [48], é interessante apresentar um exemplo simples aqui. A propriedade que falha para independência epistêmica é a seguinte:

Contração cruzada: Se X e Y são epistemicamente independentes dado Z e se X e W são epistemicamente independentes dado (Y, Z), então (Y, W) é irrelevante a X dado Z.

O seguinte exemplo mostra a falha de contração cruzada.

W	X	Y	$p_1(X,Y,W)$	$p_2(X,Y,W)$	$p_3(X,Y,W)$
W_0	X_0	Y_0	0.008	0.018	0.0093
W_1	X_0	Y_0	0.072	0.072	0.0757
W_0	X_1	Y_0	0.032	0.042	0.037
W_1	X_1	Y_0	0.288	0.168	0.228
W_0	X_0	Y_1	0.096	0.084	0.09
W_1	X_0	Y_1	0.024	0.126	0.075
W_0	X_1	Y_1	0.384	0.196	0.290
W_1	X_1	Y_1	0.096	0.294	0.195

Tabela 3.1: Densidades que definem conjunto credal no Exemplo 4.

Exemplo 4 Considere três variáveis binárias X, Y e W (neste exemplo a variável Z é omitida). Suponha que $K(W,X,Y)$ é o fecho convexo das três densidades $p_1(W,X,Y)$, $p_2(W,X,Y)$ e $p_3(W,X,Y)$ mostradas na Tabela 3.1. Este conjunto credal satisfaz independência de X e Y e também independência de X e W dado Y:

$$
\begin{aligned}
&K(X_0|Y_0) = K(X_0|Y_1) = K(X_0) \quad (p(X_0) \in [0.2, 0.3])\\
&K(Y_0|X_0) = K(Y_0|X_1) = K(Y_0) \quad (p(Y_0) \in [0.3, 0.4])\\
&K(X_0|W_0, Y_0) = K(X_0|W_1, Y_0) = K(X_0|W_0, Y_1) = K(X_0|W_1, Y_1) = K(X_0)\\
&K(W_0|X_0, Y_0) = K(W_0|X_1, Y_0) = K(W_0|Y_0) \quad (p(W_0|Y_0) \in [0.1, 0.2])\\
&K(W_0|X_0, Y_1) = K(W_0|X_1, Y_1) = K(W_0|Y_1) \quad (p(W_0|Y_1) \in [0.4, 0.8])\,.
\end{aligned}
$$

Mas para a função $h(Y,W) = I_{\{Y=Y_0, W=W_0\}}(Y,W)$, $\underline{E}[h(Y,W)|X_0] = \underline{P}(Y_0, W_0|X_0) = 0.0372$ e $\underline{E}[h(Y,W)] = \underline{P}(Y_0, W_0) = 0.04$. Consequentemente (Y,W) é relevante para X.

3.5 Condição de Kuznetsov e semi-grafóides

A condição de Kuznetsov, vista como uma definição de independência entre variáveis X e Y (condicional ou não em Z), obedece a quase todas as propriedades de grafóides, conforme mostrado no teorema a seguir. Duas observações são importantes:

- Uma suposição de não *degeneração* é feita no teorema, significando que, quando

um conjunto de faces de um conjunto credal se intersectam em um vértice, esse conjunto de faces pode ser usado para "gerar" (por combinações lineares) todos os demais hiperplanos que passam pela origem do espaço. Basicamente essa suposição requer que conjuntos credais não sejam de dimensão inferior à dimensão do espaço de distribuições (por exemplo, evitando que um conjunto credal seja um intervalo em um espaço de duas dimensões).

- A relação $X \perp\!\!\!\perp_K Y \,|\, Z$ deve ser entendida como: X e Y satisfazem a condição de Kuznetsov condicional em Z.

Teorema 7 *A condição de Kuznetsov satisfaz simetria, redundância, decomposição e união fraca, quando a aplicada a conjuntos credais não-degenerados.*

Prova. Simetria é imediata, e redundância segue de

$$\mathbf{E}[f(X)g(Y)|x_0] = f(x_0)\mathbf{E}[g(Y)|x_0] = \mathbf{E}[f(X)|x_0] \times \mathbf{E}[g(Y)|x_0]$$

para qualquer $f(X)$, $g(Y)$, e qualquer x_0. Decomposição segue do fato que qualquer função de Y é também uma função de Y and W, e portanto $\mathbf{E}[f(X)g(Y)|z] = \mathbf{E}[f(X)|z] \times \mathbf{E}[g(Y)|z]$ quando $X \perp\!\!\!\perp_K (W, Y) \,|\, Z$.

Para mostrar a propriedade de união fraca, assuma $X \perp\!\!\!\perp_K (W, Y)$; é necessário mostrar que $X \perp\!\!\!\perp_K W \,|\, Y$. Para simplificar a prova, a variável condicionante Z é omitida; essa variável pode ser adicionada a todas os valores esperados inferiores e superiores. Portanto é necessário mostrar que, para funções $f(X)$ e $g(W)$, $\underline{E}[f(X)g(W)|y] = \overline{\underline{E}}[f(X)]\overline{\underline{E}}[g(W)|y]$, onde a notação $\overline{\underline{E}}[f(X)]$ é usada para indicar "$\underline{E}[f(X)]$ ou $\overline{E}[f(X)]$, satisfazendo a condição de Kuznetsov".

Inicialmente, pode-se provar que qualquer conjunto credal satisfazendo $X \perp\!\!\!\perp_K Y \,|\, W$ tem que conter medidas de probabilidade que produzam tanto $\underline{E}[f(X)]\,\underline{E}[g(W)|y]$ quanto $\underline{E}[f(X)]\,\overline{E}[g(W)|y]$, $\overline{E}[f(X)]\,\underline{E}[g(W)|y]$ e $\overline{E}[f(X)]\,\overline{E}[g(W)|y]$. Considere a situação onde $\underline{E}[f(X)] \geq 0$ e $\underline{E}[g(W)|y] \geq 0$; nesse caso devemos ter $\underline{E}[f(X)g(W)|y] = \underline{E}[f(X)]\,\underline{E}[g(W)|y]$. Considere a função $(f(X) - \underline{E}[f(X)] + \alpha)((g(W) - \underline{E}[g(W)|y])I_y + \beta)$. A condição de Kuznetsov garante que esta função tem valor esperado inferior igual a $\alpha\beta$ para qualquer α e β; portanto deve haver uma densidade p tal que $E_p[f(X)] = \underline{E}[f(X)]$ e $E_p[g(W)|y] = \underline{E}[g(W)|y]$. Outras funções podem ser usadas para demostrar os outros casos da condição de Kuznetsov; a implicação é que sempre

existe uma densidade p (em um conjunto credal que satisfaz a condição de Kuznetsov, tal que $E_p[f(X)g(W)|y] = \underline{E}[f(X)g(W)|y]$ onde $\underline{E}[f(X)g(W)|y]$ segue a condição de Kuznetsov.

Considere agora o maior conjunto obtido a partir de $K(X)$ e $K(W,Y)$ e que satisfaz a condição de Kuznetsov $X \perp\!\!\!\perp_K (W,Y)$. Denote este conjunto por $K_k(W,X,Y)$. Se for determinado que $\underline{E}[f(X)g(W)|y] \geq \overline{\underline{E}}f(X)]\overline{\underline{E}}g(W)|y]$ para $K_k(W,X,Y)$, então automaticamente $\underline{E}[f(X)g(W)|y] = \overline{\underline{E}}f(X)]\overline{\underline{E}}g(W)|y]$ para todo conjunto satisfazendo $X \perp\!\!\!\perp_K (W,Y)$, e a propriedade de união fraca está provada. O conjunto $K_k(W,X,Y)$ satisfaz toda inequação da forma

$$\sum_{W,X,Y} f(X)h(W,Y)p(W,X,Y) \geq 0;$$

consequentemente satisfaz

$$\sum_{W,X,Y} (f(X)g(W)I_y(Y))p(W,X,Y) \geq 0$$

para toda $f(X)$, $g(W)$ e y. Se considerarmos distribuições condicionais $p(X,Y|w)$ obtidas a partir de $K_k(W,X,Y)$, essas distribuições obrigatoriamente satisfazem

$$\sum_{W,X}(f(X)g(W))p(X,W|y) \geq 0$$

(pois essa inequação é obtida normalizando a inequação anterior). Note que o maior conjunto obtido a partir de $K(X)$ e $K(W|y)$, satisfazendo $X \perp\!\!\!\perp_K W \mid y$, satisfaz inequações idênticas: $\sum_{W,X}(f(X)g(W))p(X,W|y) \geq 0$. A conclusão é que toda inequação usada para construir o maior conjunto que satisfaz $X \perp\!\!\!\perp_K W \mid y$ a partir de $K(X)$ e $K(Y|w)$ também é uma inequação satisfeita por $K_k(W,X|y)$. Portanto o primeiro conjunto é igual ou maior que o segundo. Notando que, para o primeiro conjunto temos obrigatoriamente que $\underline{E}[f(X)g(W)|y] = \overline{\underline{E}}f(X)]\overline{\underline{E}}g(W)|y]$; obtemos $\underline{E}[f(X)g(W)|y] \geq \overline{\underline{E}}f(X)]\overline{\underline{E}}g(W)|y]$ para $K_k(W,X,Y)$, e união fraca está provada. QED

A única propriedade de semi-grafóides que não é mencionada no Teorema 7 é a propriedade de contração. Como o próximo exemplo mostra, essa propriedade pode ser violada pela condição de Kuznetsov.

Exemplo 5 Considere variáveis binárias W, X, e Y, e um conjunto credal $K(W,X,Y)$ com oito vértices, tal que cada vértice se decompõe como $p(W|Y)\,p(X)\,p(Y)$. Os valores

Vértice	$p(w_0\|y_0)$	$p(w_0\|y_1)$	$p(x_0)$	$p(y_0)$
1	0.7	0.4	0.3	0.2
2	0.7	0.5	0.2	0.2
3	0.8	0.4	0.2	0.2
4	0.8	0.5	0.2	0.2
5	0.7	0.4	0.3	0.3
6	0.7	0.5	0.3	0.3
7	0.8	0.4	0.3	0.3
8	0.8	0.5	0.2	0.3

Tabela 3.2: Vértices do conjunto credal no Exemplo 5.

de $p(w_0|y_0)$, $p(w_0|y_1)$, $p(x_0)$ e $p(y_0)$ são dados na Tabela 3.2. Pode ser verificado que o conjunto credal $K(X,Y)$ satisfaz a condição de Kuznetsov (pois todas as faces desse conjunto podem ser decompostas como o produto de duas funções mais uma constante). Da mesma forma, $K(W,X|y_0)$ e $K(W,X|y_1)$ satisfazem a condição de Kuznetsov (de novo, todas as faces desses conjuntos podem ser decompostas como o produto de duas funções mais uma constante). Portanto o conjunto credal $K(W,X,Y)$ satisfaz $X \perp\!\!\!\perp_K Y$ e $X \perp\!\!\!\perp_K W \,|\, Y$; no entanto não é verdade que $X \perp\!\!\!\perp_K (W,Y)$. Tome a função $f(X) = [1,2]$ e a função $h(W,Y) = [2,1,1,2]$. Então $\underline{E}[f(X)h(W,Y)] = 2.652$ para $K(W,X,Y)$, mas $\underline{E}[f(X)]\,\underline{E}[h(W,Y)] = 1.7 \times 1.54 = 2.61$.

3.6 Discussão

Este capítulo apresentou um panorama dos conceitos de independência para conjuntos credais. A primeira conclusão importante é que os conceitos existentes não são apropriados para modelar experimentos repetitivos, situação na qual devemos recorrer ao conceito de permutabilidade. Julgamentos de permutabilidade são simples e intuitivos. Além disso, a permutabilidade em conjuntos credais tem propriedades diretamente derivadas do conceito tradicional de permutabilidade para medidas de probabilidade.

O conceito de independência é muito mais controverso que o de permutabilidade. Existem três conceitos importantes: independência forte, independência epistêmica, e a condição de Kuznetsov. O conceito de independência forte é bastante apropriado

por ser uma generalização direta do conceito tradicional de independência estocástica, e por satisfazer todas as propriedades de semi-grafóides. No entanto o conceito de independência forte requer, para sua definição, conceitos que envolvem medidas de probabilidade — uma desvantagem dentro da teoria de conjuntos credais. O conceito de independência epistêmica é simples e intuitivo, e tem uma definição diretamente ligada a conjuntos credais; no entanto, o conceito fere a propriedade de contração. A condição de Kuznetsov por sua vez é relativamente simples, embora não tão intuitiva, e também fere a propriedade de contração.

Assim sendo, a propriedade de contração parece ser um elemento fundamental na comparação entre conceitos de independência. A propriedade de contração é defendida por Pearl da seguinte forma [90, p. 85]:

> O axioma de contração estabelece que, se julgamos W irrelevante a X depois de receber alguma informação irrelevante Y, então W deve ser irrelevante [a X] antes de recebermos Y.

Nesse trecho deve-se notar primeiro que Pearl usa o termo "irrelevante", não "independente". Deve-se notar também que Pearl descreve apenas a propriedade de contração reversa. De fato, a propriedade de contração reversa é bastante desejável em um conceito de independência, e o par irrelevância/independência epistêmica tem a capacidade de representar essa propriedade de forma flexível — mais flexível até que o conceito de independência forte.

Portanto, os conceitos de irrelevância e independência epistêmicas se apresentam como os mais bem fundamentados e mais apropriados para manipulação de independência em conjuntos credais. O conceito de independência forte também apresenta excelentes características e sobretudo uma desejável semelhança com o conceito tradicional de independência estocástica. A condição de Kuznetsov não parece ter as mesmas vantagens dos outros conceitos e por isso não será abordada no restante desse trabalho.

Uma questão importante é: existe alguma maneira de unificar os conceitos de irrelevância e independência epistêmicas e o conceito de independência forte? Esta é a questão central abordada no próximo capítulo.

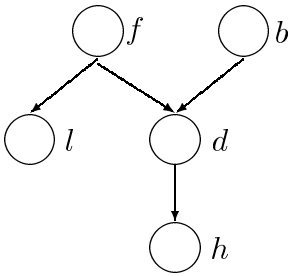
Capítulo 4

Redes credais e a condição de separação

Este capítulo foca em conjuntos credais que são representados por meio de grafos direcionados acíclicos. Tais modelos são denominados *redes credais* e tem o poder de representar várias relações de independência entre variáveis [7, 13, 58, 109]. Por isso, são modelos convenientes na investigação de conceitos de independência em conjuntos credais.

Em uma rede credal, cada nó de um grafo é associado a um conjunto credal. Há dois problemas em redes credais: primeiro, decidir qual conceito de independência deve ser utilizado em uma rede; segundo, decidir como combinar os diversos conjuntos credais associados à rede. O objetivo deste capítulo é apresentar uma teoria de redes credais que resolva essas dificuldades e leve a uma unificação dos conceitos de independência epistêmica e independência forte. Usando os conceitos do Capítulo 3, uma nova condição para independência em redes credais é proposta, denominada *condição de separação*. Esta condição de separação organiza a teoria de redes credais e conecta os diversos conceitos de independência. A partir dessa condição, redes credais passam a ter uma interpretação relativamente simples em termos de *extensões fortes*. Assim, esta dissertação atinge seu objetivo de fornecer uma teoria completa que une conceitos de independência e modelos multivariados.

O material aqui apresentado está contido em artigos publicados na *Conference on Uncertainty in Artificial Intelligence* 2000 [35] e 2002 [50].



$$
\begin{array}{ccc}
p(f) = 0.5 & p(l|f) = 0.6 & p(d|f,b) = 0.8 \\
p(b) = 0.4 & p(l|f^c) = 0.05 & p(d|f,b^c) = 0.1 \\
 & p(h|d) = 0.6 & p(d|f^c,b) = 0.1 \\
 & p(h|d^c) = 0.3 & p(d|f^c,b^c) = 0.7
\end{array}
$$

Figura 4.1: Exemplo de rede Bayesiana.

4.1 Modelos baseados em grafos direcionados acíclicos

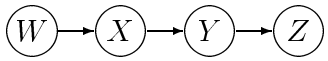
Vários modelos estatísticos multivariados podem ser representados através de grafos. As próximas seções apresentam redes Bayesianas e redes credais.

4.1.1 Redes Bayesianas

Uma rede Bayesiana é uma representação compacta e eficiente para uma medida de probabilidade. A Figura 4.1 traz um exemplo conhecido de rede Bayesiana, proposto por Charniak [14]. Este exemplo descreve uma família, envolvendo variáveis *family-out* f, *dog-health* b, *lights-on* l, *dog-out* d, e *hear-bark* h.

Uma rede Bayesiana é sempre composta por um grafo direcionado acíclico e uma coleção de variáveis $\mathbf{X}$. Cada nó do grafo é associado a uma variável X_i e com uma densidade condicional $p(X_i|\text{pa}(X_i))$ — os "pais" de X_i são indicados por $\text{pa}(X_i)$ (onde "pai" tem o sentido usado na teoria de grafos). Redes Bayesianas satisfazem a *condição de Markov*: Toda variável é independente de seus não-descendentes condicional em seus pais. A condição de Markov implica que toda rede Bayesiana representa uma densidade de probabilidade da seguinte forma:

$$p(\mathbf{X}) = \prod_i p(X_i|\text{pa}(X_i)) \,. \tag{4.1}$$



$$
\begin{array}{ll}
0.2 \leq p(w) \leq 0.3 & \\
0.1 \leq p(x|w) \leq 0.2 & 0.8 \leq p(x|w^c) \leq 0.9 \\
0.4 \leq p(y|x) \leq 0.5 & 0.5 \leq p(y|x^c) \leq 0.6 \\
0.7 \leq p(z|y) \leq 0.8 & 0.1 \leq p(z|y^c) \leq 0.2
\end{array}
$$

Figura 4.2: Exemplo de rede credal com quatro variáveis binárias.

A *moralização* de uma rede Bayesiana produz um grafo não direcionado da seguinte forma: os arcos do grafo direcionado original tem suas direções removidas, e os pais de cada variável do grafo original são ligados.

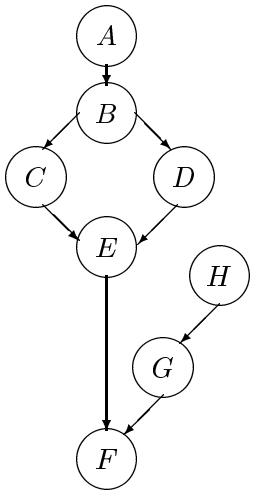
Uma das propriedades mais importantes de redes Bayesianas é a de *d-separação* [90]: se dois conjuntos de variáveis $\mathbf{W}$ e $\mathbf{Y}$ são bloqueados por um conjunto de variáveis $\mathbf{Z}$ no grafo moralizado, $\mathbf{W}$ e $\mathbf{Y}$ são estocasticamente independentes condicional on $\mathbf{Z}$. Essa propriedade é importante porque permite descobrir, em tempo polinomial ao tamanho da rede, as relações de independência estocástica na rede — com importantes implicações em algoritmos de manipulação de redes [90].

4.1.2 Redes credais

Redes credais são estruturas que representam conjuntos credais usando grafos direcionados acíclicos [28]:

Definição 4 *Uma rede credal para variáveis* $\mathbf{X}$ *é um grafo direcionado acíclico onde cada nó é associado a uma variável* X_i *em* $\mathbf{X}$ *e com um conjunto credal condicional* $K(X_i|\text{pa}(X_i))$.

Figuras 4.2 e 4.3 mostram duas redes credais. Será assumido, em todo o desenvolvimento a seguir, que os conjuntos credais associados a uma rede credal são *separadamente especificados* (Seção 2.2). Ou seja, o "conjunto credal" $K(X_i|\text{pa}(X_i))$ na verdade representa uma coleção de conjuntos credais, um para cada valor de $\text{pa}(X_i)$ — e cada um desses conjuntos é especificado independentemente dos demais. Razões



$$
\begin{array}{l}
P(a_0) \in [1/2, 3/5] \\
P(b_0|a_0) \in [1/2, 3/5] \\
P(b_0|a_1) \in [2/5, 1/2] \\
P(d_0|b_0) \in [1/5, 4/5] \\
P(d_0|b_1) \in [1/10, 1/2] \\
P(e_0|c_0, d_0) \in [1/10, 9/10] \\
P(e_0|c_0, d_1) = 1/2 \\
P(e_0|c_1, d_0) \in [3/10, 1/2] \\
P(e_0|c_1, d_1) \in [1/10, 1/2] \\
P(e_0|c_2, d_0) = 1/2 \\
P(e_0|c_2, d_1) \in [1/2, 3/5] \\
P(f_0|e_0, g_0) \in [1/5, 51/100] \\
P(f_0|e_0, g_1) \in [1/4, 1/2] \\
P(f_0|e_1, g_0) \in [9/20, 11/20] \\
P(f_0|e_1, g_1) \in [7/20, 11/20] \\
P(h_0) \in [1/5, 11/20] \\
P(g_0|h_0) \in [1/5, 1/2] \\
P(g_0|h_1) \in [3/10, 1/2]
\end{array}
$$

$$K(C|b_0) = \{[1/3, 1/3, 1/3], [7/10, 1/10, 1/5]\}$$
$$K(C|b_1) = \{[1/3, 1/3, 1/3], [1/2, 1/2, 3/10]\}$$

Figura 4.3: Exemplo de rede credal: todas as variáveis são binárias, exceto C, que tem três valores.

que justificam a restrição a conjuntos separadamente especificados estão apresentadas em [50].

Um conjunto credal $K(\mathbf{X})$ que satisfaz todas as restrições expressas por uma rede credal é uma *extensão* da rede credal. Note que a Definição 4 não prescreve uma única extensão para redes credais. Existem duas extensões que merecem destaque: *extensões fortes* e *extensões epistêmicas*. Essas extensões são estudadas nas seções seguintes.

4.2 Extensões fortes

Suponha que, a partir de uma rede credal, seja construído o maior conjunto credal para o qual toda variável é fortemente independente de seus não-descendentes conditional em seus pais. Ou seja, o conjunto credal satisfaz a condição de Markov com respeito a independência forte. Este conjunto é denominado a *extensão forte* da rede credal, e tem a seguinte forma:

Teorema 8 *A extensão forte de uma rede credal é o fecho convexo de todas as densi-*

dades de probabilidade que satisfazem a condição de Markov na rede:

$$K(\mathbf{X}) = \text{FECHO CONVEXO}\left(\begin{array}{ll} \prod_i p(X_i|\Pi_i) & : \\ p(X_i|\Pi_i) \in \boldsymbol{L}(X_i|\Pi_i) & \end{array}\right), \tag{4.2}$$

onde $\boldsymbol{L}(X_i|\text{pa}(X_i))$ *é o conjunto que contém todas as combinações permitidas de densidades de* $K(X_i|\text{pa}(X_i))$ *(o conjunto credal condicional extensivo).*

A prova deste teorema pode ser encontrada em artigo publicado pelo autor na *Conference on Uncertainty in Artificial Intelligence* 2000 [35]; o resultado aparece sem prova em várias publicações anteriores [13, 26, 58, 109].

Se toda combinação de densidades de $K(X_i|\text{pa}(X_i))$ é permitida, então $\boldsymbol{L}(X_i|\text{pa}(X_i))$ é o maior possível; nesse caso a extensão forte é dita *separadamente especificada.*

Extensões fortes tem uma similaridade intuitiva com redes Bayesianas, que se traduz no seguinte resultado:

Teorema 9 *Dado uma rede credal onde todas as combinações de variàveis tem probabilidade inferior maior que zero, toda relação de d-separação entre variáveis na rede corresponde a uma relação de independência forte na extensão forte da rede.*

A prova deste teorema também pode ser encontrada em artigo publicado pelo autor na *Conference on Uncertainty in Artificial Intelligence* 2000 [35]. O teorema demonstra que os mesmos algoritmos usados para d-separação de redes Bayesianas podem ser usados para extensões fortes.

O ponto negativo de extensões fortes é que estas extensões "herdam" os problemas de independência forte: é difícil justificar o conceito, e todas as definições existentes se baseiam em independência estocástica.

4.3 Extensões epistêmicas

Suponha que, a partir de uma rede credal, seja construído o maior conjunto credal para o qual toda variável é epistemicamente independente de seus não-descendentes conditional em seus pais. Ou seja, o conjunto credal satisfaz a condição de Markov

com respeito a independência epistêmica. Este conjunto é denominado a *extensão epistêmica* da rede credal.

Extensões epistêmicas tem uma interpretação bastante intuitiva, porém a manipulação de extensões epistêmicas é ainda um desafio. Os únicos algoritmos disponíveis para cálculo de probabilidades inferiores em extensões epistêmicas foram produzidos pelo autor em sua dissertação de doutorado [34]. A complexidade de extensões epistêmicas é enorme: a extensão epistêmica da rede na Figura 4.2 tem mais de 6 milhões de vértices! Além disso, não é sabido se extensões epistêmicas satisfazem ou não relações de d-separação. Finalmente, extensões epistêmicas levam a dificuldades na separação de sub-modelos em redes credais, como ilustrado no próximo exemplo.

Exemplo 6 Considere duas variáveis binárias X e Y tal que X e Y são epistemicamente independentes e $p(X) \in [2/5, 1/2]$, $p(Y) \in [2/5, 1/2]$; ou seja, uma rede credal trivial com dois nós desconectados. A extensão forte desta rede tem quatro pontos extremos (vetores contém valores $p(x_0, y_0)$,$p(x_0, y_1)$,$p(x_1, y_0)$, $p(x_1, y_1)$): [1/4,1/4,1/4,1/4], [4/25,6/25,6/25,9/25], [1/5,1/5,3/10,3/10],[1/5,3/10,1/5,3/10]. A extensão epistêmica tem seis pointos extremos: os quatro vértices da extensão forte e [2/9,2/9,2/9,1/3], [2/11,3/11,3/11,3/11]. Um especialista então comunica que $p(y)$ é exatamente 2/5. Para simplificar os cálculos, o especialista simplesmente calcula $p(X|Y)$ para cada vértice da extensão epistêmica e multiplica $p(X|Y)$ por 2/5 para obter novos vértices. Como X e Y são independentes, o especialista imagina que não há necessidade de mudar a probabilidade condicional $p(X|Y)$ e recalcular toda a extensão epistêmica. Realizando esses cálculos, o especialista obtém quatro pontos: [4/25,6/25,6/25,9/25], [1/5,3/10,1/5,3/10], [1/5,6/25,1/5,9/25], [4/25,3/10,6/25,3/10]. Porém, nesse novo modelo X e Y *não* são epistemicamente independentes! Por exemplo, $p(y_0) = 2/5$ mas $p(y_0|x_0) \in [2/5, 4/7]$.

Outro exemplo, bastante informal, é o seguinte. Tome três variáveis, X como {ensolarado, nublado}, Y como {aeroporto funcionando, aeroporto fechado}, W as {bom tempo, mal tempo}. Suponha que temos a rede credal da Figura 4.2 (sem a variável Z), e independência epistêmica entre Y e W condicional em X. Nesse modelo, uma modificação na probabilidade de W pode levar a mudanças na probabilidade de Y condicional em X — se essas mudanças não forem feitas, as variáveis podem violar independência epistêmica. No entanto, parece absurdo mudar o modelo para W apenas porque houve uma mudança em uma variável que é "independente" de W.

4.4 A condição de separação

As duas seções anteriores podem ser resumidas como segue: extensões fortes são algoritmicamente similares a redes Bayesianas, mas enfrentam os problemas de interpretação do conceito de independência forte; extensões epistêmicas são intuitivas mas não satisfazem algumas condições básicas que deveriam ser esperadas de um conceito de "independência". O problema é que independência epistêmica não tem comportamento desejável quando associada à condição de Markov. A seguinte questão pode ser formulada: existe algum tipo de modelo para redes credais que, partindo de independência epistêmica, tem as mesmas propriedades de separação que redes Bayesianas?

A idéia central proposta nesse trabalho é que a condição de Markov tradicional é muito fraca quando aplicada a redes credais. Para obter uma condição apropriada a redes credais, considere a seguinte interpretação para a expressão "mudanças na crença". Suponha que um conjunto credal é dado. Todo vértice de $K(X,Y)$ pode ser escrito como $p(X|Y)\,p(Y)$. Selecione um vértice $p(X,Y)$, remova-o da coleção de pontos extremos de $K(X,Y)$ e modifique-o: substitua $p(Y)$ por uma densidade arbitrária $p'(Y)$. Então insira $p(X|Y)\,p'(Y)$ de novo na coleção de pontos extremos de $K(X,Y)$ e tome o fecho convexo. Essa modificação será denominada uma *mudança de crença em $K(X,Y)$ com respeito a Y*.

Considere agora a seguinte condição:

Definição 5 (Condição de separação) *Toda variável X_i é epistemicamente independente de seus não-descendentes condicional em seus pais, mesmo que haja uma mudança de crença em $K(X,Y)$ com respeito a X_i.*

O próximo teorema demostra que a condição de separação atinge seus objetivos: quando imposta a uma rede credal, a condição de separação leva a uma extensão que satisfaz as condições de separação em uma rede Bayesianas — de fato, a condição de separação leva a uma extensão forte!

Teorema 10 *Em uma rede credal, a condição de separação é válida se e somente se toda variável X_i é fortemente independente de seus não-descendentes condicional em seus pais.*

4.5 Discussão

Para entender as consequências da condição de separação na teoria de redes credais, note que os Teoremas 8 e 10 implicam a equivalência de

1. Condição de separação (com independência epistêmica); e
2. Extensão forte.

Portanto, independência epistêmica satisfaz as relações de separação de redes Bayesianas, desde que associada à condição de separação. Além disso, extensões fortes não dependem diretamente do conceito de independência forte, mas sim do conceito de independência epistêmica. Independência forte pode ser vista como um conceito derivado de independência epistêmica e condição de separação.

Como discutido no Capítulo 3, independência forte enfrenta dois problemas. Em primeiro lugar, depende da referência a densidades dentro de um conjunto credal. Em segundo lugar, depende da definição de independência estocástica. O desenvolvimento apresentado nesse trabalho, culminando na condição de separação, elimina a necessidade de independência estocástica. Também é possível eliminar a necessidade de mencionar densidades; para isso, é necessário definir *mudança de crença* de forma a não mencionar densidades. Isso é possível: basta dizer que a mudança de crença com respeito a Y ocorre pela adição de restrições de probabilidade por valores esperados inferiores (note que $\underline{E}[f(X)] = \alpha$ equivale a $E_p[f(X)] \geq \alpha$). Isso leva a uma situação em que a teoria de conjuntos credais pode ser inteiramente desenvolvida sem menção a conceitos particulares da teoria de medidas de probabilidade.

A manipulação de extensões fortes oferece grandes desafios computacionais. Extensões fortes podem ter um enorme número de vértices, e o cálculo de valores esperados inferiores a partir de extensões fortes leva a problemas NP-difíceis mesmo em redes cujo grafo é uma árvore [50].

Capítulo 5

Conclusão

Este trabalho procurou descrever as atividades de pesquisa do autor no período 1998-2002, dando ênfase á sistematização de um particular tópico de pesquisa: conceitos de independência em conjuntos credais. O Capítulo 1 procurou traçar um panorama geral das atividades do autor, enquanto os demais capítulos focaram em conjuntos credais.

A teoria de conjuntos credais é apropriada em uma série de aplicações relacionadas a tomada de decisões — situações onde a teoria convencional de probabilidades enfrenta dificuldades. O Capítulo 2 procurou mostrar quando a teoria é recomendada, as razões do interesse despertado pela teoria na área de inteligência artifical, e sobretudo procurou introduzir os conceitos básicos utilizados na teoria. O Capítulo 2 enfatizou aspectos algorítmicos de conjuntos credais, procurando mostrar o estado-da-arte para manipulação de modelos que não contém julgamentos de independência. Boa parte dos algoritmos mais eficientes conhecidos foi desenvolvida pelo autor.

O Capítulo 3 é voltado para a análise dos conceitos de independência e permutabilidade propostos na literatura. Através da análise de exemplos e do estudo das propriedades de semi-grafóides, os conceitos de independência forte e independência epistêmica se mostraram bastante satisfatórios. Porém o conceito de independência forte demonstra certa fragilidade quanto à sua motivação e justificativa. Esse problema se estende a modelos baseados em grafos direcionados e independência forte — as extensões fortes de redes credais. O conceito de independência epistêmica é mais intuitivo mas também enfrenta dificuldades quando aplicado a redes credais. O autor tem contribuído regularmente em relação a esses temas, e os resultados descritos resumem

significativa porção dos resultados existentes sobre conjuntos credais e conceitos de independência.

O Capítulo 4 descreve as dificuldades relacionadas a redes credais e propõe uma solução, na forma da condição de separação. Em resumo, tomando independência epistêmica como a definição de independência para conjuntos credais, temos:

$$\text{Condição de separação} \Leftrightarrow \text{Extensão forte.} \tag{5.1}$$

A equivalência expressa em (5.1) lembra a relação existente em redes Bayesianas, onde:

$$\text{Condição de Markov} \Leftrightarrow \text{Expressão (4.1).}$$

Em última análise, a equivalência (5.1) é a razão pela qual a condição de separação pode ser uma base sólida para a teoria de redes credais.

O trabalho aqui descrito fornece meios para que a teoria de redes credais possa ser explorada de forma organizada. Uma série de desafios de pesquisa pode ser abordada a partir dos resultados aqui apresentados. Um dos maiores desafios é desenvolver algoritmos que possam manipular com eficiência as diversas extensões de redes credais. Outro desafio é generalizar a teoria para casos limite, como situações em que conjuntos credais são degenerados ou eventos condicionantes tem probabilidade inferior igual a zero. Finalmente, o maior desafio é provavelmente aplicar a teoria de conjuntos credais a problemas práticos que possam revelar os pontos fortes e os pontos fracos da teoria.

Referências Bibliográficas

[1] M. Ackermann e F. G. Cozman. Um sistema de travamento para a articulação do joelho de órteses de membros inferiores. *II Congresso Nacional de Engenharia Mecânica*, págs. 116–124, João Pessoa, Paraíba, Brasil, 2002.

[2] M. Ackermann, F. G. Cozman, e A. S. Dias. Sistema para flexão de joelho mecanizada em órteses de membros inferiores. *Anais do XVI Congresso Brasileiro de Engenharia Mecânica*, volume 1 (Bioengenharia), pág. 254–263. Associação Brasileira de Ciências Mecânicas, 2001.

[3] J. C. Adamowski, M. G. Simões, e F. G. Cozman. Desenvolvimento de um robô móvel. *VIII Congresso Brasileiro de Automática*, 1990.

[4] J. C. Adamowski, C. M. Furukawa, e F. G. Cozman. Mecatrônica — uma abordagem voltada à automação industrial. *XXVIII Congresso Brasileiro de Ensino de Engenharia*, Ouro Preto, Minas Gerais, Brasil, 2000.

[5] I. Blavatsky-Magalhães e F. G. Cozman. Aprendizado ativo utilizando dados classificados e não-classificados. *XV Simpósio Nacional de Probabilidade e Estatística*, pág. 190, Lindóia, São Paulo, Brasil, 2002.

[6] I. Blavatsky-Magalhães e F. G. Cozman. Avaliando o erro de classificação em aprendizado de máquina semi-supervisionado via bootstrap. *XV Simpósio Nacional de Probabilidade e Estatística*, pág. 191, Lindóia, São Paulo, Brasil, 2002.

[7] K. A. Andersen e J. N. Hooker. Bayesian logic. *Decision Support Systems*, 11:191–210, 1994.

[8] F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge: A Logical Approach*. MIT Press, Cambridge, 1990.

[9] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[10] J. O. Berger. Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25:303–328, 1990.

[11] J. S. Breese e K. W. Fertig. Decision making with interval influence diagrams. P. P. Bonissone, M. Henrion, L. N. Kanal, e J. F. Lemmer, editores, *Uncertainty in Artificial Intelligence 6*, pág. 467–478. Elsevier Science, North-Holland, 1991.

[12] G. Bruno e A. Gilio. Applicazione del metodo del simplesso al teorema fondamentale per le probabilità nella concezione soggettivistica. *Statistica*, 40:337–344, 1980.

[13] J. Cano, M. Delgado, e S. Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8:253–280, 1993.

[14] E. Charniak. Bayesian networks without tears. *AI Magazine*, págs. 50–63, 1991.

[15] L. Chrisman. Incremental conditioning of lower and upper probabilities. *International Journal of Approximate Reasoning*, 13(1):1–25, 1995.

[16] L. Chrisman. Propagation of 2-monotone lower probabilities on an undirected graph. E. Horvitz e F. Jensen, editores, *XII Uncertainty in Artificial Intelligence Conference*, págs. 178–186, San Francisco, Califórnia, Estados Unidos, 1996. Morgan Kaufmann.

[17] I. Cohen, A. Bronstein, e F. G. Cozman. Adaptive online learning of Bayesian network parameters. Technical Report HPL-2001-156, Hewlett-Packard Laboratories, Palo Alto, Califórnia, 2001.

[18] G. Coletti e R. Scozzafava. Coherent upper and lower Bayesian updating. G. de Cooman, F. G. Cozman, S. Moral, e P. Walley, editores, *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, págs. 101–110, Ghent, Bélgica, 1999. Imprecise Probabilities Project, Universiteit Gent.

[19] V. M. H. Coupé, L. L. C. van der Gaag, e J. D. F. Habbema. Sensitivity analysis: an aid for belief-network quantification. *Knowledge Engineering Review*, 15:1–18, 2000.

[20] I. Couso, S. Moral, e P. Walley. Examples of independence for imprecise probabilities. G. de Cooman, F. G. Cozman, S. Moral, e P. Walley, editores, *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, págs. 121–130, Ghent, Bélgica, 1999. Imprecise Probabilities Project, Universiteit Gent.

[21] I. Couso, S. Moral, e P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5:165–181, 2000.

[22] F. G. Cozman e E. Krotkov. Robot localization using a computer vision sextant. *International Conference on Robotics and Automation*, págs. 106–111, 1995.

[23] F. G. Cozman e E. Krotkov. Quasi-Bayesian strategies for efficient plan generation: Application to the planning to observe problem. *XII Uncertainty in Artificial Intelligence Conference*, págs. 186–193, 1996.

[24] F. G. Cozman e E. Krotkov. Automatic mountain detection and pose estimation for teleoperation of lunar rovers. *Proc. of the International Conference on Robotics and Automation*, págs. 2452–2457, 1997.

[25] F. G. Cozman e E. Krotkov. Depth from scattering. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, págs. 801–806, 1997.

[26] F. G. Cozman. Robustness analysis of Bayesian networks with local convex sets of distributions. D. Geiger e P. Shenoy, editores, *XIII Uncertainty in Artificial Intelligence Conference*, págs. 108–115, San Francisco, 1997. Morgan Kaufmann.

[27] F. G. Cozman. Independence relations in the robustness analysis of multivariate probabilistic models. *XII Conferência Brasileira de Automática*, V:1891–1896, 1998.

[28] F. G. Cozman. Irrelevance and independence relations in Quasi-Bayesian networks. G. Cooper e S. Moral, editores, *XIV Conference on Uncertainty in Artificial Intelligence*, págs 89–96, San Francisco, 1998. Morgan Kaufmann.

[29] F. G. Cozman e E. Krotkov. Position estimation from outdoor visual landmarks for teleoperation of lunar rovers. *Proc. Third IEEE Workshop on Applications of Computer Vision*, págs. 156–161, 1996.

[30] F. G. Cozman e P. E. Miyagi. Trajectory controller for a mobile robot using optimal control. *XI Congresso Brasileiro de Engenharia Mecânica*, 3:537–540, 1991.

[31] F. G. Cozman. Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. *Journal of Computational and Graphical Statistics*, 8(4):824–838, 1999.

[32] F. G. Cozman. Irrelevance and independence axioms in quasi-Bayesian theory. Anthony Hunter e Simon Parsons, editores, *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, págs. 128–136. Springer, London, England, 1999.

[33] F. G. Cozman. Computing posterior upper expectations. *International Journal of Approximate Reasoning*, 24:191–205, 2000.

[34] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[35] F. G. Cozman. Separation properties of sets of probabilities. C. Boutilier e M. Goldszmidt, editores, *XVI Conference on Uncertainty in Artificial Intelligence*, págs. 107–115, San Francisco, 2000. Morgan Kaufmann.

[36] F. G. Cozman. Independence relations in the robustness analysis of multivariate probabilistic models. *Revista Ciência e Engenharia*, 1:30–36, 1999.

[37] F. G. Cozman. Sensitivity and robustness analysis of Bayesian networks. *IV Simpósio Brasileiro de Automação Industrial*, págs. 251–255, 1999.

[38] F. G. Cozman. Generalizing variable elimination in Bayesian networks. L. N. de Barros, R. Marcondes, F. G. Cozman, e A. H. R. Costa, editores, *Workshop on Probabilistic Reasoning in Artificial Intelligence*, págs. 27–32, São Paulo, 2000. Editora Tec Art.

[39] F. G. Cozman. The JavaBayes system. *The ISBA Bulletin*, 7(4):16–21, 2000.

[40] F. G. Cozman. Removing redundancy in the expectation-maximization algorithm for Bayesian network learning. L. N. de Barros, R. Marcondes, F. G. Cozman, e A. H. R. Costa, editores, *Workshop on Probabilistic Reasoning in Artificial Intelligence*, págs. 21–26, São Paulo, 2000. Editora Tec Art.

[41] F. G. Cozman. Constructing sets of probability measures through Kuznetsov's independence condition. *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, págs. 104–111, Ithaca, New York, 2001.

[42] F. G. Cozman. Algorithms for conditioning on events of zero lower probability. *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, págs. 248–252, Pensacola, Flórida, 2002.

[43] F. G. Cozman e I. Cohen. Unlabeled data can degrade classification performance of generative classifiers. Technical Report HPL-2001-234, Hewlett-Packard Laboratories, Palo Alto, Califórnia, 2001.

[44] F. G. Cozman e I. Cohen. Unlabeled data can degrade classification performance of generative classifiers. *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, págs. 327–331, Pensacola, Flórida, 2002.

[45] F. G. Cozman, J. C. F. da Rocha, e C. P. de Campos. Computing with sets of probability measures. *Workshop on Validated Computing*, page 45, Toronto, Canadá, 2002.

[46] F. G. Cozman e C. M. Furukawa. A reestruturação curricular do curso de Mecatrônica da Escola Politécnica. *XXVIII Congresso Brasileiro de Ensino de Engenharia*, Ouro Preto, Minas Gerais, Brasil, 2000.

[47] F. G. Cozman, E. Krotkov, e C. E. Guestrin. Outdoor visual position estimation for planetary rovers. *Autonomous Robots*, 9:135–150, 2000.

[48] F. G. Cozman e P. Walley. Graphoid properties of epistemic irrelevance and independence. *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, págs. 112–121, Ithaca, New York, 2001.

[49] J. C. F. da Rocha e F. G. Cozman. Evidence propagation in credal networks: an exact algorithm based on separately specified sets of probability. *Brazilian Symposium on Artificial Intelligence (SBIA)*, págs. 376–385, Porto de Galinhas, Pernambuco, Brasil, 2002.

[50] J. C. F. da Rocha e F. G. Cozman. Inference with separately specified sets of probabilities in credal networks. A. Darwiche e N. Friedman, editores, *Proceedings of the XVIII Conference on Uncertainty in Artificial Intelligence*, págs. 430–437, San Francisco, Califórnia, 2002. Morgan Kaufmann.

[51] A. P. Dawid. Conditional independence. S. Kotz, C. B. Read, e D. L. Banks, editores, *Encyclopedia of Statistical Sciences, Update Volume 2*, págs. 146–153. Wiley, New York, 1999.

[52] L. de Campos e S. Moral. Independence concepts for convex sets of probabilities. P. Besnard e S. Hanks, editores, *XI Conference on Uncertainty in Artificial Intelligence*, págs. 108–115, San Francisco, Califórnia, 1995. Morgan Kaufmann.

[53] G. de Cooman, F. G. Cozman, S. Moral, e P. Walley, editores. *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications.* Imprecise Probabilities Project, Universiteit Gent, Ghent, Bélgica, 1999.

[54] B. de Finetti. *Theory of probability, vol. 1-2.* Wiley, New York, 1974.

[55] L. DeRobertis e J. A. Hartigan. Bayesian inference using intervals of measures. *Annals of Statistics*, 9(2):235–244, 1981.

[56] D. Dubois, H. Prade, e P. Smets. Representing partial ignorance. *IEEE Transactions on Systems, Man and Cybernetics A*, 26(3):361–377, 1996.

[57] R. Fagin, J. Y. Halpern, e N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990.

[58] E. Fagiuoli e M. Zaffalon. 2U: An exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.

[59] T. L. Fine. Lower probability models for uncertainty and nondeterministic processes. *Journal of Statistical Planning and Inference*, 20:389–411, 1988.

[60] A. M. Frisch e P. Haddawy. Anytime deduction for probabilistic logic. *Artificial Intelligence*, 69:93–122, 1994.

[61] D. Galles e J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97:9–43, 1997.

[62] D. Geiger, T. Verma, e J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.

[63] F. J. Giron e S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? J. M. Bernardo, J. H. DeGroot, D. V. Lindley, e A. F. M. Smith, editores, *Bayesian Statistics*, págs. 17–38. University Press, Valencia, Espanha, 1980.

[64] I. J. Good. *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press, Minneapolis, 1983.

[65] P. Gritzmann e V. Klee. Mathematical programming and convex geometry. P. Gruber e J. M. Wills, editores, *Handbook of Convex Geometry*, volume A, págs. 627–674. Elsevier Science Publishers, 1993.

[66] C. E. Guestrin, F. G. Cozman, e E. Krotkov. Fast software image stabilization with color registration. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, págs. 19–24, 1998.

[67] H. E. Kyburg Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[68] T. Hailperin. *Sentential Probability Logic*. Lehigh University Press, Bethlehem, Estados Unidos, 1996.

[69] J. Y. Halpern e R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[70] P. Hansen, B. Jaumard, M. P. de Aragão, F. Chauny, e S. Perron. Probabilistic satisfiability with imprecise probabilities. *International Journal of Approximate Reasoning*, 24(2-3):171–189, 2000.

[71] T. Herron, T. Seidenfeld, e L. Wasserman. Divisive conditioning: Further results on dilation. Technical Report 585, Department of Statistics, Carnegie Mellon University, 1993.

[72] J. S Ide e F. G. Cozman. Raciocínio probabilístico em sistemas embarcados. *Anais do XXI Congresso da Sociedade Brasileira de Computação*, pág. 220. Sociedade Brasileira de Computação, 2001.

[73] J. S. Ide e F. G. Cozman. Geração de redes Bayesianas por Metropolis-Hastings. *XV Simpósio Nacional de Probabilidade e Estatística*, page 193, Lindóia, São Paulo, Brasil, 2002.

[74] J. S. Ide e F. G. Cozman. Random generation of Bayesian networks. *Brazilian Symposium on Artificial Intelligence (SBIA)*, págs. 366–375, Porto de Galinhas, Pernambuco, Brasil, 2002.

[75] J. S. Ide e F. G. Cozman. Testing MCMC algorithms with randomly generated Bayesian networks. *I Workshop de Teses e Dissertações em Inteligência Artificial*, págs. 252–261, Porto de Galinhas, Pernambuco, Brasil, 2002.

[76] B. Jaumard, P. Hansen, e M. P. de Aragão. Column generation methods for probabilistic logic. *ORSA Journal on Computing*, 3(2):135–148, Spring 1991.

[77] H. Kyburg Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, 1961.

[78] J. B. Kadane, editor. *Robustness of Bayesian Analyses*, Elsevier Science Pub. Co., New York, 1984.

[79] V. P. Kuznetsov. *Interval Statistical Methods*. Radio i Svyaz Publ., 1991.

[80] K. B. Laskey. Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(6):901–909, 1995.

[81] M. Lavine. Sensitivity in Bayesian statistics, the prior and the likelihood. *Journal of the American Statistical Association*, 86(414):396–399, 1991.

[82] M. Lavine, L. Wasserman, e R. L. Wolpert. Bayesian inference with specified prior marginals. *Journal of the American Statistical Association*, 86(416):964–971, 1991.

[83] J. F. Lemmer e H. E. Kyburg Jr. Conditions for the existence of belief functions corresponding to intervals of belief. *9th National Conference on Artificial Intelligence*, págs. 488–493, 1991.

[84] I. Levi. On indeterminate probabilities. *Journal of Philosophy*, 71:391–418, 1874.

[85] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts, 1980.

[86] T. Lukasiewicz. Uncertain reasoning in concept lattices. *3rd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, págs. 293–300, 1995.

[87] C. Luo, C. Yu, J. Lobo, G. Wang, e T. Pham. Computation of best bounds of probabilities from uncertain data. *Computational Intelligence*, 12(4):541–566, 1996.

[88] N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[89] M. P. Pacifico, G. Salinetti, e L. Tardella. Fractional optimization in Bayesian robustness. Technical Report Serie A n. 23, Dipartamento di Statistica, Probabilita e Statistiche Applicate, Universita di Roma La Sapienza, Italy, 1994.

[90] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, Califórnia, 1988.

[91] L. R. Pericchi e M. E. Perez. Posterior robustness with more than one sampling model. *Journal of Statistical Planning and Inference*, 40:279–294, 1994.

[92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, e B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridgeshire, 1992.

[93] F. T. Ramos e F. G. Cozman. Embedded Bayesian networks: Anyspace, anytime probabilistic inference. *Workshop on Real-Time Decision Support and Diagnosis Systems*, págs. 13–19, Edmonton, Canadá, 2002.

[94] F. T. Ramos e F. G. Cozman. Selecting exact and approximate algorithms in decomposed Bayesian networks. *Workshop de Computação*, págs. 151–155, São José dos Campos, São Paulo, Brasil, 2002.

[95] F. T. Ramos, F. Mikami, e F. G. Cozman. Implementação de redes Bayesianas em sistemas embarcados. *Proceedings of the IBERAMIA/SBIA 2000 Workshops (Workshop on Probabilistic Reasoning in Artificial Intelligence)*, págs. 65–69. Editora Tec Art, 2000.

[96] E. H. Ruspini. The logical foundations of evidential reasoning. Technical Report SRIN408, SRI International, 1987.

[97] S. J. Russell e P. Norvig. *Artificial Intelligence: a Modern Approach.* Upper Saddle River, Prentice Hall, New Jersey, 1995.

[98] S. I. Schaible e W. T. Ziemba. *Generalized Concavity in Optimization and Economics.* Academic Press, 1981.

[99] T. Seidenfeld e M. Schervish. Two perspectives on consensus for (Bayesian) inference and decisions. *IEEE Transactions on Systems, Man and Cybernetics*, 20(1):318–325, 1990.

[100] T. Seidenfeld, M. J. Schervish, e J. B. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23(6):2168–2217, 1995.

[101] T. Seidenfeld e L. Wasserman. Dilation for sets of probabilities. *Annals of Statistics*, 21(9):1139–1154, 1993.

[102] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[103] E. H. Shortliffe e B. G. Buchanan. *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project.* The Addison-Wesley series in artificial intelligence. Addison-Wesley, Reading, Mass., 1985.

[104] R. Simmons, E. Krotkov, L. Chrisman, F. Cozman, R. Goodwin, M. Hebert, L. Katragadda, S. Koenig, G. Krishnaswamy, Y. Shinoda, W. Whittaker, e P. Klarer. Experience with rover navigation for lunar-like terrains. *Proc. Intl. Conf. Intelligent Robots and Systems (IROS)*, págs. 441–446, Pittsburgh, Pennsylvania, 1995.

[105] C. A. B. Smith. Consistency in statistical inference and decision. *Journal Royal Statistical Society B*, 23:1–25, 1961.

[106] P. Snow. Improved posterior probability estimates from prior and conditional linear constraint systems. *Transactions on Systems, Man, and Cybernetics A*, 21(2):464–469, 1991.

[107] M. Studeny. Semigraphoids and structures of probabilistic conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21(1):71–98, 1997.

[108] P. Suppes. The measurement of belief. *Journal Royal Statistical Society B*, 2:160–191, 1974.

[109] B. Tessem. Interval probability propagation. *International Journal of Approximate Reasoning*, 7:95–120, 1992.

[110] H. Thone, U. Guntzer, e W. Kieβling. Towards precision of probabilistic bounds propagation. *VIII Uncertainty in Artificial Intelligence Conference*, págs. 315–322, 1992.

[111] P. Walley. Coherent lower (and upper) probabilities. Technical Report Statistics Report 23, University of Warwick, Coventry, 1981.

[112] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[113] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

[114] P. Walley, R. Pelessoni e Paolo Vicig. Direct algorithms for checking coherence and making inferences from conditional probability assessments. Dipartimento di Matematica Applicata B. de Finetti, Università di Trieste, Trieste, Itália, 1999.

[115] P. Walley e T. L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10(3):741–761, 1982.

[116] L. Wasserman. Recent methodological advances in robust Bayesian inference. J. M. Bernardo, J. O. Berger, A. P. Dawid, e A. F. M. Smith, editores, *Bayesian Statistics 4*, págs. 483–502. Oxford University Press, 1992.

[117] L. A. Wasserman e J. B. Kadane. Bayes' theorem for Choquet capacities. *Annals of Statistics*, 18(3):1328–1339, 1990.

[118] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149–170, 2000.

[119] C. C. White III. A posteriori representations based on linear inequality descriptions of a priori and conditional probabilities. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-16(4):570–573, 1986.