

MATHIAS JUAN PERAZZO

DERIVADAS EM REDES BAYESIANAS USANDO
ELIMINAÇÃO DE VARIÁVEIS

Dissertação Apresentada à Escola Politécnica
da Universidade de São Paulo para Obtenção
do Título de Mestre em Engenharia.

São Paulo

2005

MATHIAS JUAN PERAZZO

DERIVADAS EM REDES BAYESIANAS USANDO
ELIMINAÇÃO DE VARIÁVEIS

Dissertação Apresentada à Escola Politécnica
da Universidade de São Paulo para Obtenção
do Título de Mestre em Engenharia

Área de Concentração: Inteligência Artificial

Orientador:

Prof. Dr.

Fabio G. Cozman

São Paulo

2005

Aos meus pais, Alberto e Beatriz, a
meu irmão Marcos e a Claudia.

Agradecimentos

Ao meu amigo e orientador, Prof. Dr. Fabio Gagliardi Cozman, por todo suporte e orientação dados durante os três anos de mestrado.

Ao Prof. Dr. Gilberto Francisco Martha de Souza e ao Prof. Dr. Newton Maruyama pelas sugestões recebidas no Exame de Qualificação.

À Prof. Dr. Veerle Coupé pelo auxílio prestado durante o início do trabalho

Ao amigo e Mestre Fábio Tozeto Ramos por me indicar ao nosso orientador.

Ao amigo e companheiro de congressos Mestre Eric Conrado de Souza.

Ao amigo Thiago de Castro Martins não só pelo companheirismo nos congressos, mas também pelo auxílio prestado em diversas ocasiões.

Ao colega e Mestre André Saheki do Laboratório de Tomada de Decisão da POLI.

E, especialmente, ao amigo e Mestre Jaime Shinsuke Ide pelo suporte logístico na POLI, sem o qual seria impossível completar este trabalho.

Resumo

Redes Bayesianas são extensivamente usadas em inteligência artificial, reconhecimento de padrões e identificação de sistemas. Uma operação importante é a diferenciação de redes Bayesianas; isto é, o cálculo de derivadas de parâmetros de uma rede Bayesiana. Este trabalho apresenta um novo método para este cálculo: apresentamos um algoritmo baseado no método de eliminação de variáveis, e discutimos as vantagens e aplicações deste novo algoritmo.

Abstract

Bayesian networks are extensively used in artificial intelligence, pattern recognition and system identification. An important operation is the differentiation of Bayesian networks; that is, the computation of derivatives for the parameters of a Bayesian network. This work presents a new method for such a computation: we present an algorithm based on the variable elimination method, and discuss the advantages and applications of this new algorithm.

Sumário

Lista de Figuras	iii
Lista de Tabelas	vi
1 INTRODUÇÃO	1
2 REDES BAYESIANAS	4
2.1 Teoria básica dos grafos	4
2.2 Redes Bayesianas	6
2.3 Árvores de junção (<i>Junction Trees</i>)	8
2.4 Algoritmos para inferência em redes Bayesianas	10
2.5 Eliminação de variáveis em redes Bayesianas	11
2.6 Generalizando eliminação de variáveis	14
3 DIFERENCIAÇÃO DE REDES BAYESIANAS	17
3.1 Variação sistemática	17
3.2 Quociente de duas funções lineares	18
3.2.1 Probabilidades sem influência	18
3.2.2 Relações de funções	22
3.3 Método proposto por Kjaerulff e van der Gaag	26
3.4 Comparação dos métodos exatos	28
3.5 Redes Bayesianas com abordagem diferencial	30
3.6 Aplicações	30
3.6.1 Análise de sensibilidade de redes Bayesianas	31
3.6.2 Otimização local de parâmetros	32

4	UM NOVO MÉTODO BASEADO EM ELIMINAÇÃO DE VARIÁVEIS	34
4.1	Algoritmo proposto	35
4.2	Comparação do método proposto com métodos existentes	36
4.3	Demonstração do método	38
4.4	Conclusões	43
5	RESULTADOS EXPERIMENTAIS	44
5.1	Programa desenvolvido em Java	45
5.2	Rede de exemplo (5 nós)	46
5.3	Rede Alarm (37 nós)	47
5.4	Rede Cardiomiopatia (28 nós)	54
6	CONCLUSÕES	61
	Referências Bibliográficas	63
	APÊNDICES	66
A	ARQUIVOS DE ENTRADA	67
A.1	Rede de exemplo (5 nós)	67
A.2	Rede Alarm (37 nós)	68
A.3	Rede Cardiomiopatia (28 nós)	79

Lista de Figuras

2.1	Tipos de estruturas de grafos empregados em redes Bayesianas [17].	5
2.2	Exemplo de rede Bayesiana.	7
2.3	Transformação de um GDA em uma árvore de junção.	9
2.4	Exemplo de uma árvore de junção.	10
2.5	Árvore de <i>buckets</i> e seus separadores para a rede de exemplo (E como evidência).	14
3.1	$Insen_1$	20
3.2	$Insen_2$	20
3.3	$Insen_3$	21
3.4	$Insen_3$	21
3.5	Exemplo de rede.	23
3.6	Procedimento para refinamento das probabilidades.	32
4.1	Comparação do novo método proposto com o método proposto por Kjaerulff e van der Gaag.	37
4.2	Transformando em árvore de <i>Buckets</i>	39
4.3	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(e \neg c)$	41
4.4	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(c \neg a)$	42
4.5	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(a)$	42
5.1	Rede de exemplo.	46

5.2	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(\neg a)$	47
5.3	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(c a)$	48
5.4	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(\neg c a)$	48
5.5	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(\neg c \neg a)$	49
5.6	Curva de derivada: probabilidade de interesse $p(c e)$ e probabilidade em estudo $p(e c)$	49
5.7	Rede Alarm.	50
5.8	Rede Alarm - conjunto de sensibilidade.	51
5.9	Curva de derivada: probabilidade de interesse $p(CO = Low BP = Low)$ e probabilidade em estudo $p(Hypovolemia = True)$	51
5.10	Curva de derivada: probabilidade de interesse $p(CO = Low BP = Low)$ e probabilidade em estudo $p(StrokeVolume = Low LVFailure = False, Hypovolemia = False)$	52
5.11	Curva de derivada: probabilidade de interesse $p(CO = Low BP = Low)$ e probabilidade em estudo $p(Catechol = High InsuffAnesth = False, SaO2 = Low, TPR = Low, ArtCO2 = Normal)$	53
5.12	Rede Cardiomiopatia.	54
5.13	Rede Cardiomiopatia - conjunto de sensibilidade.	55
5.14	Curva de derivada: probabilidade de interesse $p(Ins.Card.=Sim Evidências)$ e probabilidade em estudo $p(Febre Reumática=Não)$	56
5.15	Curva de derivada: probabilidade de interesse $p(Ins.Card.=Sim Evidências)$ e probabilidade em estudo $p(Idade=66-75)$	56
5.16	Rede Cardiomiopatia 2.	57
5.17	Rede Cardiomiopatia - conjunto de sensibilidade 2.	58
5.18	Curva de derivada: probabilidade de interesse $p(Ins.Card.=Sim Evidências)$ e probabilidade em estudo $p(Hipertensão Pulmonar Venosa Sistêmica=Não Valvopatia=Não, Insuficiência Cardíaca=Não, Nível Ativação N-H=Não)$	59

5.19 Curva de derivada: probabilidade de interesse $p(\text{Ins.Card.}=\text{Sim}|\text{Evidências})$
e probabilidade em estudo $p(\text{Doença Chagas}=\text{Não})$ 60

Lista de Tabelas

3.1	Tabela comparando os métodos de derivadas exatos.	29
4.1	Tabela dos valores das probabilidades.	38
5.1	Tabela de coeficientes da rede de exemplo.	46
5.2	Tabela de coeficientes das três probabilidades em estudo.	50
5.3	Tabela de coeficientes das duas probabilidades em estudo.	55
5.4	Tabela de coeficientes das duas probabilidades em estudo 2.	58

Capítulo 1

INTRODUÇÃO

Esta dissertação tem como foco a diferenciação de redes Bayesianas; isto é, o cálculo de derivadas de parâmetros de uma rede Bayesiana.

Redes Bayesianas são modelos voltados à tomada de decisão e à manipulação de incertezas, com aplicações em inteligência artificial, reconhecimento de padrões, identificação de sistemas [18]. São modelos gerais e compactos, com capacidade para representar situações estáticas e dinâmicas.

Redes Bayesianas são utilizadas como base de conhecimento em sistemas especialistas. Sistemas especialistas com redes Bayesianas são aplicados em diversas áreas: diagnóstico, prognóstico e planejamento do tratamento médico [7, 1]; suporte técnico para solucionadores de problemas (*troubleshooters*) [15]; sistemas de decisão para interpretar dados de telemetria em tempo real [16]; genética [13]; reconhecimento de fala [30]; interpolação automática de dados brutos de sondas interplanetárias [28]; sistemas especialistas como o Pathfinder e o Munin [18], etc.

Derivadas de parâmetros de redes Bayesianas são de grande utilidade em pelo menos duas aplicações:

1. *Realização de análise de sensibilidade em redes Bayesianas.* Em uma rede Bayesiana nem todo valor de probabilidade vai requerer o mesmo nível de precisão para que a inferência na rede tenha um resultado satisfatório; algumas probabilidades têm um impacto maior na rede do que outras. Dessa forma, quando se deseja refinar os valores de probabilidades (determinar o nível de precisão requerido para as diversas probabilidades condicionais da rede), é recomendado realizar uma análise de sensibilidade [7, 5]. Numa análise de

sensibilidade deriva-se a rede Bayesiana: uma derivada onde a saída é o valor da probabilidade de interesse e se está derivando em relação aos demais valores de probabilidade da rede. Através desta derivada descobrem-se quais são as variáveis que exercem maior influência, valores de probabilidades para os quais o sistema mostrou ser mais sensível. O refinamento (ou otimização) pode ser focado nessas probabilidades.

2. *Otimização de parâmetros por métodos de gradiente.* Em muitas ocasiões o valor exato de uma probabilidade de uma rede Bayesiana pode não ser conhecido. Isso ocorre quando o especialista construindo a rede tem informações vagas ou insuficientes; ou quando um conjunto de especialistas diverge sobre os parâmetros de uma rede; ou quando dados experimentais são coletados e não são suficientes para estimar um único valor de probabilidades. Nesses casos, uma solução possível é considerar intervalos de probabilidades e realizar inferências a partir desses intervalos [3, 9, 14]. A realização de inferências requer então a otimização de valores de probabilidades com as restrições prescritas por intervalos. Em tais casos, um método de inferência importante é a otimização local por método do gradiente, a qual exige o cálculo de derivadas em cada uma de suas iterações. Os resultados apresentados nesse trabalho podem então ser utilizados. Otimização local pode também ser aplicada em estimação de parâmetros (por exemplo máxima verossimilhança).

Um dos objetivos deste trabalho é apresentar um método exato para se obter derivadas em redes Bayesianas baseado no *algoritmo de eliminação de variáveis*. Esse método é de simples implementação computacional e de fácil aprendizado, pois diferente dos algoritmos existentes (Capítulo 3), não envolve conceitos gráficos, tais como triangulação e cliques, evitando dessa forma a utilização de complexos métodos da teoria de grafos. Ao invés disso, o algoritmo proposto se foca apenas em densidades de probabilidade.

Outro objetivo é implementar esse novo método e integrá-lo ao sistema JavaBayes. JavaBayes é um pacote em Java livremente distribuído para a construção e manipulação de redes Bayesianas ¹. Esse pacote é desenvolvido sob a orientação do

¹JavaBayes está disponível em <http://www.pmr.poli.usp.br/ltd/Software/javabayes/>.

pesquisador Fábio G. Cozman (que, não por coincidência, também é o orientador deste trabalho) no Laboratório de Tomada de Decisão da Escola Politécnica.

O Capítulo 2 contém um breve resumo da teoria de redes Bayesianas e do algoritmo de eliminação de variáveis aplicado em redes Bayesianas. O Capítulo 3 por sua vez faz uma análise dos métodos existentes para cálculo de derivadas em redes Bayesianas e indica duas possíveis aplicações para a teoria aqui apresentada, e o Capítulo 4 apresenta o método proposto neste trabalho incluindo um exemplo resolvido passo a passo pelo método proposto. O Capítulo 5 descreve o programa em Java desenvolvido e apresenta os resultados experimentais obtidos com três redes. Finalmente o Capítulo 6 conclui o trabalho realizado e apresenta os próximos passos.

Capítulo 2

REDES BAYESIANAS

Este capítulo tem como objetivo apresentar os conceitos que serão utilizados ao longo do trabalho. Nas Seções 2.1, 2.2 e 2.3 apresenta-se um breve resumo sobre teoria dos grafos, redes Bayesianas e árvores de junção respectivamente. Na Seção 2.4 são citados os principais métodos de inferência, exatos e aproximados, em redes Bayesianas. Dentre esses métodos de inferência, o método de Eliminação de Variáveis e sua generalização são descritos com mais detalhe nas Seções 2.5 e 2.6.

2.1 Teoria básica dos grafos

Esta seção apresenta as principais definições da teoria de grafos. Maiores detalhes sobre teoria de grafos são encontrados em [6, 4].

A facilidade na visualização e na construção de modelos são as duas principais vantagens em se utilizar grafos na representação de conhecimento.

Um *grafo* é denotado por um par $G = (\mathbf{V}, \mathbf{E})$, onde \mathbf{V} é um conjunto finito de *vértices*, também conhecidos como nós, e \mathbf{E} é um subconjunto do conjunto $\mathbf{V} \times \mathbf{V}$ de pares ordenados de vértices, denominados arcos de G . O arco (u, v) é direcionado se $(u, v) \in \mathbf{E}$ mas $(v, u) \notin \mathbf{E}$. Num arco direcionado (u, v) , u é um pai de v . Se todos os arcos forem direcionados, então o *grafo* G é um *grafo direcionado*, senão o *grafo* G é um *grafo não-direcionado*. Dois nós u e v são *adjacentes* se o arco (u, v) pertence ao conjunto \mathbf{E} . Um *caminho* é uma seqüência de nós de modo que cada par de nós consecutivos são adjacentes. Um caminho é um *ciclo* se contém o mesmo nó como ponto inicial e final. Um ciclo é *direcionado* se é possível alcançar um mesmo

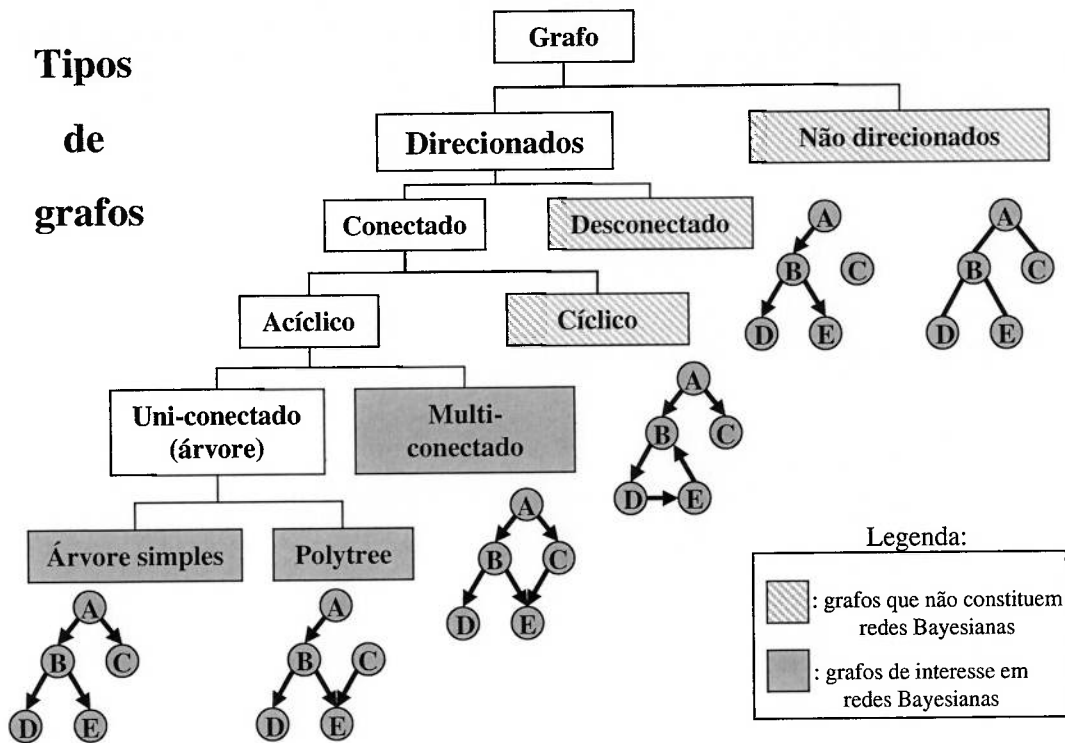


Figura 2.1. Tipos de estruturas de grafos empregados em redes Bayesianas [17].

nó seguindo arcos que estão na mesma direção. Um grafo é chamado *acíclico*, se não possui nenhum ciclo. Um *grafo moral* é um grafo não-direcionado gerado a partir de um grafo direcionado com a conexão dos pais de cada nó v_i e removendo as direções dos arcos. Um grafo moral é *triangularizado* se, e somente se, em todo ciclo de comprimento quatro ou maior que quatro o grafo possui um arco que conecta dois nós não-adjacentes [25]. Cliques em um grafo triangularizado são conjuntos de variáveis onde cada variável do conjunto é conectada às outras. Um grafo é *conectado* se existe caminho entre quaisquer dois nós pertencentes ao conjunto de vértices V . Um *grafo direcionado acíclico* (GDA) é um grafo direcionado que não contém nenhum ciclo direcionado. Para todo nó v contido em V , o *grau* é o número de arcos contendo v . Em outras palavras, *grau* é o número total de arcos entrando ou saindo do nó v [17].

Árvore é um grafo em que existe exatamente um caminho entre quaisquer dois nós do grafo. Se existir mais de um caminho o grafo será *multi-conectado*. A árvore será uma *árvore simples* se ela possuir apenas uma *raiz* (um nó é *raiz*, se não possui nenhum pai). Caso contrário ela será *polytree*.

Um resumo da classificação dos tipos de grafos, segundo características previa-

mente descritas, é apresentado na Figura 2.1. Os GDAs (árvores simples, polytrees e grafos multi-conectados) são utilizados em redes Bayesianas. Os demais grafos (grafos não direcionados, desconectados e cíclicos) não constituem redes Bayesianas.

2.2 Redes Bayesianas

Assume-se neste trabalho que todas as variáveis aleatórias têm um número finito de possíveis valores. Um conjunto de variáveis é escrito em negrito (por exemplo, \mathbf{X}). A notação $\mathbf{X} \setminus \mathbf{Y}$ indica o conjunto de variáveis que pertence a \mathbf{X} mas não a \mathbf{Y} . A notação $\sum_{\mathbf{X}} f(\mathbf{X}, \mathbf{Y})$ indica que todas as variáveis em \mathbf{X} são eliminadas (somando) da função $f(\mathbf{X}, \mathbf{Y})$. A densidade de probabilidade de X é $p(X)$: $p(x)$ é o valor da probabilidade do evento $\{X = x\}$. A densidade de probabilidade de X condicional a valores de Y é $p(X|Y)$.

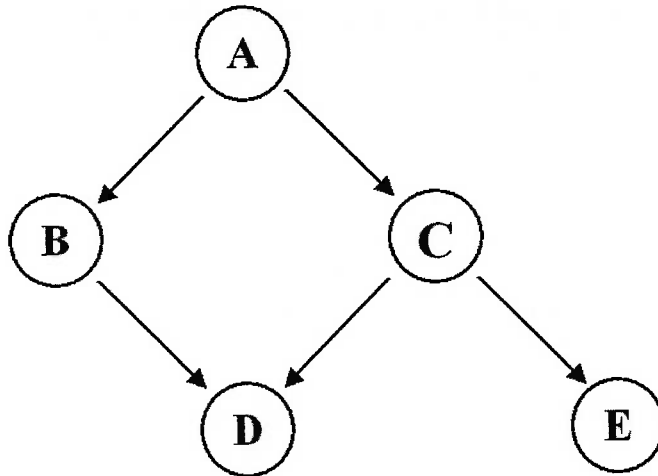
Uma rede Bayesiana representa uma densidade de probabilidade sobre um conjunto de variáveis \mathbf{X} [23, 18], disposta em um grafo direcionado acíclico (GDA) com n nós.

Uma rede Bayesiana é portanto composta de duas partes:

- Parte qualitativa: Grafo direcionado acíclico composto pelos nós (variáveis no domínio) e pelos arcos (relacionamento de influência entre os nós).
- Parte quantitativa: Distribuições de probabilidades condicionais associadas a cada um dos nós.

A densidade é especificada através do grafo direcional acíclico. Cada nó nesse grafo representa uma variável aleatória X_i em \mathbf{X} . O conjunto de variáveis $pa(X_i)$ denota os pais de X_i , $ch(X_i)$ denota os filhos de X_i . O conjunto composto por X_i e todos os seus predecessores é denotado por $\pi(X_i)$ e $\sigma(X_i)$ denota o conjunto composto por X_i e todos os seus sucessores. Os pais dos filhos de X_i que não sejam eles próprios filhos são as “esposas” $spo(X_i)$. A família de um nó X_i é o conjunto formado pela união de X_i com seus pais $pa(X_i)$. Cada variável X_i é associada a uma distribuição $p(X_i | pa(X_i))$ sendo que uma variável sem pais é associada à distribuição marginal $p(X_i)$. A Figura 2.2 ilustra relações em uma rede.

A propriedade básica em redes Bayesiana é que elas atendem à condição de Markov: cada nó é condicionalmente independente de seus não descendentes não pais



Probabilidades na rede:

$$p(A), p(B|A), p(C|A), p(D|B, C), p(E|C)$$

Figura 2.2. Exemplo de rede Bayesiana.

dados seus pais. Essa condição leva a uma distribuição conjunta de probabilidade única [23]:

$$p(\mathbf{X}) = \prod_i p(X_i | pa(X_i)).$$

Dada uma rede Bayesiana, o evento E denota as *evidências* ou *observações* na rede. Por exemplo $E = \{X_1 = x_{12}, X_3 = x_{31}\}$ fixa os valores das variáveis X_1 e X_3 .

Inferência com redes Bayesianas envolve normalmente o cálculo da probabilidade marginal *a posteriori* para um conjunto de variáveis questionadas \mathbf{X}_q [23]. A probabilidade *a posteriori* de \mathbf{X}_q dado E é:

$$P(\mathbf{X}_q | E) = \frac{P(\mathbf{X}_q, E)}{P(E)} = \frac{\sum_{\mathbf{X} \setminus \{\mathbf{X}_q, \mathbf{X}_E\}} P(\mathbf{X})}{\sum_{\mathbf{X} \setminus \mathbf{X}_E} P(\mathbf{X})}.$$

Computacionalmente o termo importante é o numerador:

$$P(\mathbf{X}_q, E) = \propto \sum_{\mathbf{X} \setminus \{\mathbf{X}_q, \mathbf{X}_E\}} P(\mathbf{X}),$$

já que uma vez obtido o numerador, pode-se facilmente obter o denominador. O denominador $p(E)$ é simplesmente uma constante de normalização já que a evidência E é fixa [10].

Outro conceito em redes Bayesianas é d-separação:

d-separação [23] *Dado X, Y, Z , suponha que por todo caminho de uma variável em X até uma variável em Y , existe uma variável W tal que:*

- *W tem arcos convergentes, não está em Z nem seus descendentes;*
- *W não está em Z .*

Então, X e Y estão d-separados por Z ; $\langle X|Z|Y \rangle^d$

Como d-separação é bastante utilizado ao longo deste trabalho, este conceito está descrito em mais detalhes na subseção 3.2.1.

2.3 Árvores de junção (*Junction Trees*)

Alguns algoritmos de inferência probabilística, tais como *clustering*¹ ou eliminação, utilizam uma estrutura auxiliar para representar o conhecimento em uma etapa intermediária no processo de inferência [25]. Esta estrutura é conhecida como *árvore de junção*.

Neste trabalho um dos métodos de diferenciação (descrito na Seção 3.3) é construído em cima de redes Bayesianas na sua representação em árvores de junção. Como será visto mais adiante esse método constitui uma das bases (a outra base é o Método de Eliminação de Variáveis Estendido descrito na Seção 2.5) para a proposição do novo método de diferenciação de redes Bayesianas.

A propriedade fundamental de árvores de junção é [8]:

Definição *Considere C como sendo uma coleção de sub-conjuntos de um conjunto finito V de variáveis de um GDA, e τ uma árvore com C sendo seu conjunto de nós. Então, τ é uma árvore de junção se satisfizer a seguinte propriedade:*

- *Qualquer intersecção $C_1 \cap C_2$ de um par de sub-conjuntos C_1, C_2 em C tem de estar contida em todo nó pertencente ao caminho único entre C_1 e C_2 . Simultaneamente, cada nó v em V terá sua família incluída em, no mínimo, um sub-conjunto.*

¹Foram mantidos neste trabalho os termos em inglês, *cluster* e *bucket* a fim de evitar confusões com as referências utilizadas neste trabalho, em particular [7, 21, 10].

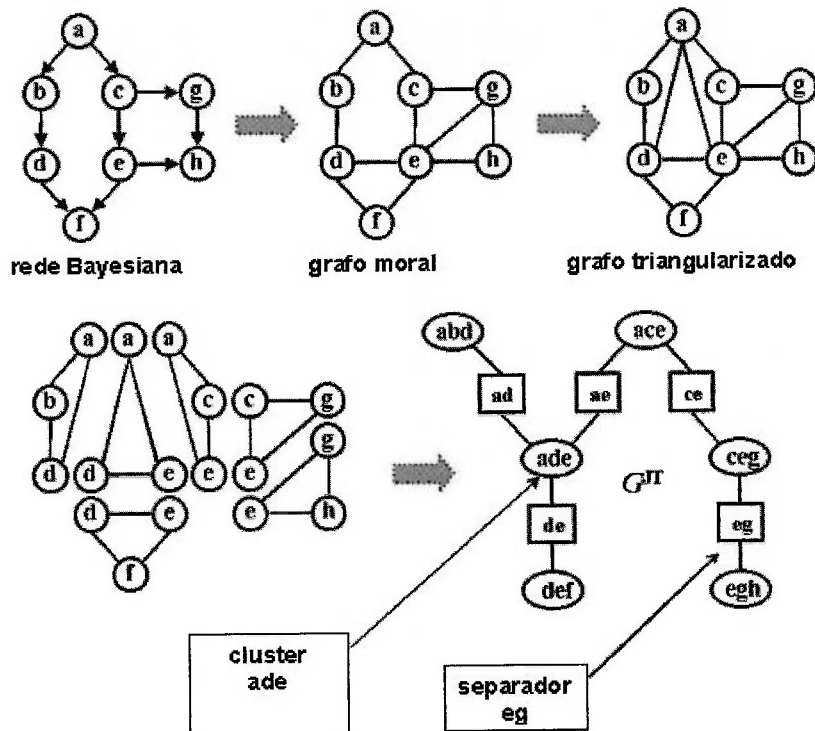


Figura 2.3. Transformação de um GDA em uma árvore de junção.

Numa árvore de junção os nós C_i são definidos como *clusters* e as variáveis associadas com a intersecção $C_1 \cap C_2$ são definidas como *separadores*.

Transformar um GDA em uma árvores de junção envolve, geralmente, as seguintes etapas [11]:

1. Construção de um grafo moral a partir do GDA;
2. Adição de arcos ao grafo moral, a fim de deixá-lo triangularizado;
3. A partir do grafo triangularizado, identificar os cliques e associá-los a *clusters*;
4. Construção da árvore de junção conectando-se os *clusters* de forma a obter, ao final do processo, uma árvore não-direcionada que satisfaça a propriedade fundamental das árvores de junção.

Na Figura 2.3 pode-se ver tais etapas. Mais detalhes sobre árvores de junção e sobre sua construção estão disponíveis em [19, 20].

A Figura 2.4 mostra a árvore de junção obtida a partir da rede de exemplo 2.2 seguindo as quatro etapas de construção. A árvore possui três *clusters* (ABC , BCD e CE) e dois *separadores* (BC e C). Pode-se observar também que a árvore de junção

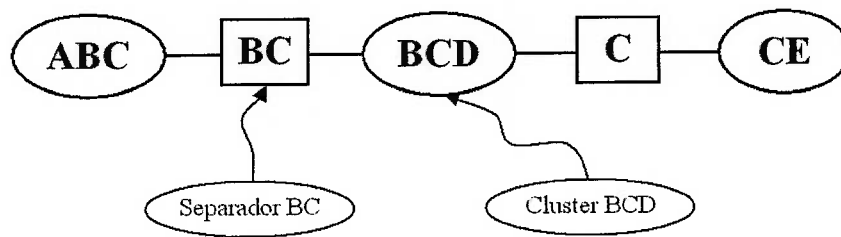


Figura 2.4. Exemplo de uma árvore de junção.

satisfaz a propriedade de sua definição: o nó C que é a intersecção entre os *clusters* ABC e CE está contido em todo o caminho único entre ABC e CE , assim como o nó C tem sua família incluída no *cluster* ABC .

2.4 Algoritmos para inferência em redes Bayesianas

Os algoritmos para inferência em redes Bayesianas podem ser de dois tipos: inferência exata ou inferência aproximada.

Existem vários algoritmos para inferência exata em redes Bayesianas. Eles podem basicamente ser divididos em três categorias:

- Propagação de mensagens em *Polytree*.
- Método do Condicionamento (*cutset conditioning*).
- Eliminação de Variáveis e *Clustering*.

Algoritmos de inferência exata podem ser exponenciais no tamanho do maior clique existente no grafo da rede triangulado [12]. Isso faz com que a inferência exata de muitos modelos de interesse seja muito lenta. Uma alternativa é recorrer às técnicas da aproximação [26].

Este trabalho foca no método de **eliminação de variáveis**. O método de eliminação de variáveis, na sua essência, simplesmente elimina uma a uma as variáveis através da soma. Dada a sua importância para este trabalho, este método será descrito em mais detalhes na Seção 2.5.

2.5 Eliminação de variáveis em redes Bayesianas

Eliminação de variáveis é um algoritmo para inferência exata em redes Bayesianas [12, 29].

Dada uma rede Bayesiana de variáveis \mathbf{X} , evidência E e variáveis questionadas \mathbf{X}_q , inferir $p(\mathbf{X}_q|E)$ envolve normalmente apenas um subconjunto das densidades associadas com a rede. Se a densidade $p(X_i|pa(X_i))$ for necessária para produzir uma inferência então X_i é uma *variável requisitada*; existem algoritmos polinomiais para encontrar todas as variáveis requisitadas para uma inferência. O conjunto de variáveis requisitadas é expresso por \mathbf{X}_R . Todas as variáveis em \mathbf{X}_q pertencem à \mathbf{X}_R , mas nem todas as variáveis observadas pertencem à \mathbf{X}_R . Apenas variáveis observadas que possuem pais não observados em \mathbf{X}_R pertencem a \mathbf{X}_R (a definição de variáveis requisitadas será vista em detalhes no capítulo seguinte).

Deseja-se calcular:

$$p(\mathbf{X}_q, \mathbf{E}) = \sum_{\mathbf{X}_R \setminus \{X_q, X_E\}} \left(\prod_{X_i \in \mathbf{X}_R} p(X_i|pa(X_i)) \right) \quad (2.1)$$

onde as densidades de probabilidades estão restritas a domínios que não contém evidência.

Seja N o número de variáveis requisitadas que não são observadas e não estão em \mathbf{X}_q . Suponha que essas variáveis estejam ordenadas de uma determinada maneira² em que se tenha $\{X_1, X_2, X_3, \dots, X_N\}$. Então:

$$p(\mathbf{X}_q, \mathbf{E}) = \sum_{X_1 \dots X_N} \left(p(X_N|pa(X_N)) \times \dots \times p(X_1|pa(X_1)) \right), \quad (2.2)$$

que pode ser reescrita da seguinte forma:

²A ordem das variáveis é arbitrária, mas diferentes ordenações levam a diferentes cargas computacionais. Infelizmente, a construção de um ordenador ótimo é um problema NP-completo, então deve-se aceitar uma ordenação boa, não necessariamente ótima. Existem vários métodos heurísticos conhecidos que produzem ordenações eficientes [29, 20].

$$p(\mathbf{X}_q, \mathbf{E}) = \sum_{X_N} \cdots \sum_{X_1} p(X_N | pa(X_N)) \times \cdots \times p(X_1 | pa(X_1)). \quad (2.3)$$

A variável X_1 pode apenas aparecer em densidades $p(X_j | pa(X_j))$ para $X_j \in \{X_1, ch(X_1)\}$. Pode-se então mover a somatória de X_1 :

$$p(\mathbf{X}_q, \mathbf{E}) = \sum_{X_N} \cdots \sum_{X_2} \left(\prod_{\substack{X_i \in \\ X_R \setminus \{X_1, ch(X_1)\}}} p(X_i | pa(X_i)) \right) \\ \times \left(\sum_{X_1} \prod_{\substack{X_j \in \\ \{X_1, ch(X_1)\}}} p(X_j | pa(X_j)) \right),$$

sendo que se pode definir a seguinte densidade não normalizada:

$$p(ch(X_1) | pa(X_1), spo(X_1)) = \sum_{X_1} \left(\prod_{X_j \in \{X_1, ch(X_1)\}} p(X_j | pa(X_j)) \right).$$

Ou seja, pode-se eliminar a variável X_1 do problema. Repetindo-se sucessivamente para X_2 até X_N no final teremos algumas (pelo menos uma) densidades para \mathbf{X}_q . Através da multiplicação dessas densidades e normalizando o resultado obtemos $p(\mathbf{X}_q | E)$.

Para facilitar o entendimento das equações acima, usa-se a rede Bayesiana da Figura 2.2 como exemplo. Tendo $\{C = c\}$ como variável questionada e $\{E = e\}$ como evidência, através de d-separação determina-se que A , C e E são variáveis requisitadas. Da Equação 2.1:

$$p(c, e) = \sum_{X_R \setminus \{X_q, X_E\}} \left(\prod_{X_i \in X_R} p(A) p(c | A) p(e | c) \right),$$

que pode ser reescrito como:

$$p(c, e) = \sum_{X_A} \left(\prod_{X_i \in X_R} p(A) p(c | A) p(e | c) \right).$$

Eliminado-se a variável requisitada e a evidência das variáveis requisitadas sobrou apenas a variável A . Com apenas uma variável não é preciso se preocupar com o ordenamento.

$$p(c, e) = \left(p(a)p(c|a)p(e|c) + p(\neg a)p(c|\neg a)p(e|c) \right)$$

Assumindo os valores $p(a) = 0,20$, $p(c|a) = 0,20$, $p(c|\neg a) = 0,05$, $p(e|c) = 0,80$:

$$p(c, e) = 0,064.$$

Utilizando-se os mesmos princípios tem-se:

$$p(\neg c, e) = 0,552.$$

Normalizando obtém-se: $c = 0,103896$ e $\neg c = 0,896104$.

O algoritmo de eliminação de variáveis pode ser descrito da seguinte forma [10]:

Algoritmo *Eliminação de Variáveis*

1. Gere um ordenamento para as N variáveis que são necessárias para a inferência, que não são observadas e não questionadas.
2. Coloque todas as densidades da rede em um conjunto de densidades.
3. Para i de 1 a N :
 - a. Crie uma estrutura de dados B_i , chamada *bucket*, contendo:
 - A variável X_i , chamada variável do *bucket*;
 - Todas as densidades que contenham a variável do *bucket*, chamado de densidade do *bucket*, mas que ainda não foram colocadas em nenhum *bucket*;
 - b. Multiplique as densidades em B_i . Armazene a densidade resultante não normalizada em B_i ; a densidade é chamada de *cluster* de B_i .
 - c. Elimine (somando) X_i do *cluster* B_i . Armazene a densidade resultante não normalizada em B_i ; a densidade é chamada de *separador* de B_i .
 - d. Coloque o separador do *bucket* no conjunto de densidades.
4. No final do processo colete as densidades que contém a variável questionadas num *bucket* B_q . Multiplique as densidades em B_q juntas e normalize o resultado.

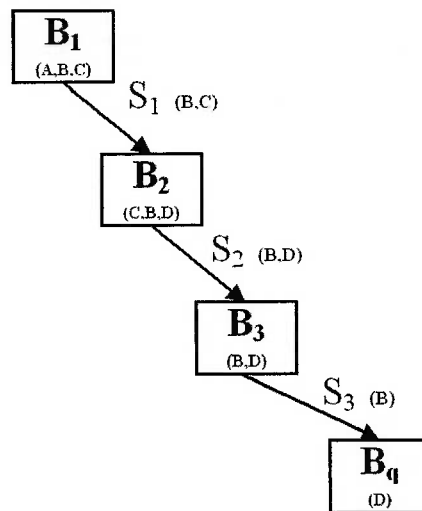


Figura 2.5. Árvore de *buckets* e seus separadores para a rede de exemplo (E como evidência).

Uma vez que a ordenação das variáveis esteja definida, o algoritmo de eliminação de variáveis cria estruturas de dados chamadas *buckets*. Cada *bucket* contém uma variável a ser eliminada X_i e todas as densidades que contenham a variável do *bucket*. A seqüência de *buckets* criada é chamada de *árvore de buckets*. Cada *bucket* (um de cada vez) é processado através da coleta de densidades da rede Bayesiana, computando o separador do *bucket* e enviando esse separador ao longo da árvore de *buckets*.

2.6 Generalizando eliminação de variáveis

Suponha agora que se deseje computar a densidade de probabilidade marginal para cada variável numa rede Bayesiana. Através do armazenamento de resultados intermediários no processo de eliminação, pode-se produzir inferências para todas as variáveis na rede de forma conjunta [10].

Considerando B_a como o *bucket* logo acima da raiz e S_a como seu separador, para atualizar esse *bucket*, temos [10]:

$$p(X_a, \mathbf{X}_q | E) = p(X_a | \mathbf{S}_a, E_a) \times p(\mathbf{X}_q | E), \quad (2.4)$$

sendo que a densidade atualizada X_a é então obtida eliminando X_q (através da soma) da densidade $p(X_a, \mathbf{X}_q|E)$. Dessa forma atualiza-se todos os *buckets* que estão logo acima da raiz.

Considere agora um *bucket* B_b a alguma distância da raiz. Assuma que eliminação de variáveis já foi executada e que B_b esteja normalizado em relação à X_b ; tem-se então que o *cluster* B_b contém a densidade normalizada $p(X_b|\mathbf{S}_b, E_b)$.

Suponha que B_c , um *cluster* logo abaixo de B_b , já foi atualizado, ou seja, B_c contém a densidade $p(X_c, \mathbf{S}_c|E)$. \mathbf{S}_b tem que estar contido nas variáveis do *cluster* B_c (chamadas de C_c), pois o separador \mathbf{S}_b foi enviado à B_c durante a eliminação de variáveis. Pode-se dessa forma calcular em B_c

$$p(\mathbf{S}_b|E) = \sum_{C_c \setminus \mathbf{S}_b} p(C_c|E). \quad (2.5)$$

Pode-se então calcular:

$$p(X_b, \mathbf{S}_b|E) = p(X_b|\mathbf{S}_b, E_b) \times p(\mathbf{S}_b|E). \quad (2.6)$$

Finalmente a densidade atualizada X_b é então obtida eliminando \mathbf{S}_b (através da soma) da densidade $p(X_b, \mathbf{S}_b|E)$.

Considerando que o filho do *bucket* B_i já foi atualizado, o algoritmo que generaliza de Eliminação de Variáveis pode ser descrito da seguinte forma [10]:

Algoritmo *Eliminação de Variáveis Estendido*

1. Normalize o *cluster* de B_i em relação à variável do *bucket*, obtendo-se a densidade normalizada $p(X_i|\mathbf{S}_i, E_i)$.
2. Peça ao filho de B_i que forneça $p(\mathbf{S}_i|E_i)$. Esta densidade de probabilidade substitui o separador de B_i .
3. Multiplique $p(\mathbf{S}_i|E_i)$ e o *cluster* normalizado do *bucket* $p(X_i|\mathbf{S}_i, E_i)$, obtendo $p(X_i, \mathbf{S}_i|E_i)$. Esta densidade de probabilidade substitui o *cluster* de B_i .

A propagação na generalização parte da raiz em direção às extremidades. Após atualização completa cada *bucket* conterá $p(X_i, \mathbf{S}_i | E)$, possibilitando então através de uma simples soma obter a densidade de probabilidade para qualquer variável X_i .

Capítulo 3

DIFERENCIAÇÃO DE REDES BAYESIANAS

Numa diferenciação de rede Bayesiana deseja-se investigar os efeitos causados em uma dada probabilidade de interesse pela variação do valor de uma ou mais probabilidades condicionais da parte quantitativa da rede.

Existem vários métodos para diferenciar redes Bayesianas; alguns métodos são exatos e outros são aproximados. A seguir serão citados alguns dos métodos mais importantes.

Quando um método, algoritmo ou definição for relevante para o trabalho, esse tópico será descrito em mais detalhes. Um tópico é considerado relevante quando este for utilizado ou servir de base para alguma parte do novo método proposto neste trabalho (descrito no Capítulo 4).

3.1 Variação sistemática

Esta primeira abordagem não é um método propriamente dito, pois não calcula exatamente derivadas, mas produz indiretamente o seu valor. Cada parâmetro de probabilidade da rede é sistematicamente variado e o efeito na probabilidade de saída (probabilidade de *interesse*) é investigado. Obtém-se dessa forma a variação em relação à condição inicial. Esse método tem a desvantagem de consumir muito esforço computacional, pois cada variação em valores deve ser acompanhada de uma inferência.

3.2 Quociente de duas funções lineares

O método proposto por Coupé *et al* [7] requer consideravelmente menos esforços computacionais do que simplesmente variar sistematicamente as probabilidades condicionais. É um método construído na sua maioria sobre a parte qualitativa da rede. Este método tem dois princípios:

- Como o diagrama de uma rede representa as independências ao longo das variáveis estatísticas envolvidas, ele permite identificar probabilidades condicionais cujas variações não podem influenciar a probabilidade de interesse. Análises a respeito dessas probabilidades condicionais não são informativas e podem portanto, ser excluídas da análise.
- As análises que são informativas podem ser representadas por simples expressões matemáticas. Mais especificamente, a probabilidade de interesse de uma rede Bayesiana pode ser expressa como um quociente de duas funções lineares de uma probabilidade condicional em estudo. As constantes dessa função determinam a sensibilidade da probabilidade de interesse à probabilidade condicional em estudo. Determinar essas constantes requer apenas algumas inferências na rede.

Como esses princípios também serão usados no método proposto neste trabalho, eles serão descritos em mais detalhes nas subseções seguintes.

3.2.1 Probabilidades sem influência

Sabe-se de antemão que para várias probabilidades condicionais de uma rede Bayesiana, suas variações não afetam a probabilidade de interesse. Essas probabilidades estão *d-separadas* da probabilidade de interesse e podem ser prontamente identificadas através de inspeções no diagrama da rede, sem que seja necessário nenhum cálculo probabilístico. Uma derivada de rede Bayesiana pode se restringir às probabilidades condicionais que afetam a probabilidade de interesse (que não estejam *d-separadas* da probabilidade de interesse). Os nós aos quais essas probabilidades se referem constituem o conjunto de sensibilidade (*sensitivity set*) para o nó de interesse.

Definição formal do conjunto de sensibilidade [7]:

Definição Seja B uma rede Bayesiana com o grafo $G = (V(G), A(G))$. Seja $V_r \in V(G)$ o nó de interesse e seja $O \subseteq V(G)$ o conjunto de nós observados em G . Seja G^* o grafo que é construído de G pela adição de um predecessor auxiliar X_i a cada nó $V_i \in V(G)$. Então, o conjunto de sensibilidade para V_r dado O , chamado $Sen(V_r, O)$, é o conjunto de todos os nós $V_i \in V(G)$ para os quais X_i e V_r não são d-separados em G^* .

O predecessor auxiliar X_i do nó V_i , citado na definição, pode ser interpretado como representação para imprecisão no valor da probabilidade para V_i . Se a presença de imprecisão no valor de V_i não for d-separada do nó de interesse, ou seja, se V_r não for isolada da imprecisão pela evidência disponível, então a variação do valor da probabilidade condicional de V_i pode influenciar o valor da probabilidade de V_r . Portanto V_r faz parte do conjunto de sensibilidade.

Uma abordagem mais fácil do que determinar as probabilidades que fazem parte do conjunto de sensibilidade é determinar as que não fazem. A seguir são definidos três conjuntos de nós que englobam todas as possíveis probabilidades condicionais que não têm influência (estão d-separadas) na probabilidade de interesse da rede [7]:

Definição Seja B uma rede Bayesiana com o grafo G . Seja $V_r \in V(G)$ o nó de interesse e seja O o conjunto de nós observados, como na definição anterior. Defina-se os conjuntos de nós $Insen_1(V_r, O)$, $Insen_2(V_r, O)$ e $Insen_3(V_r, O)$, respectivamente como:

- para cada nó $V_i \in \pi(V_r)$, se $\langle (\{V_i\} \cup pa(V_i)) | O | \{V_r\} \rangle^d$, então $V_i \in Insen_1(V_r, O)$;
- para cada nó $V_i \in V(G) \setminus \pi(V_r)$, se $\langle (\{V_i\} \cup pa(V_i)) | O | \{V_r\} \rangle^d$ e $\sigma(V_i) \cap O \neq \emptyset$, então $V_i \in Insen_2(V_r, O)$;
- para cada nó $V_i \in V(G) \setminus \pi(V_r)$, se $\sigma(V_i) \cap O = \emptyset$, então $V_i \in Insen_3(V_r, O)$;

Essa segunda definição está determinando quais são os nós que estão d-separados do nó de interesse V_r . A princípio a definição acima pode parecer mais complexa do que a definição direta do conjunto de sensibilidade, mas as quatro figuras seguintes indicam claramente o significado e a simplicidade da definição dos conjuntos $Insen_3$.

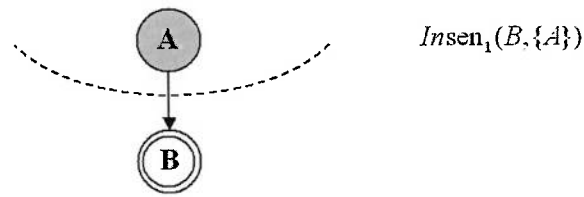


Figura 3.1. $Insen_1$.

Na rede de apenas dois nós acima, sendo B a probabilidade de interesse, o conjunto de nós $Insen_1$ é formado pelo nó A , que é uma evidência. Ou seja, variação nos valores de probabilidades do nó A e de qualquer nó que viesse a existir como ascendente de A não influenciaria o valor da probabilidade de interesse B , pois o nó pai A é uma evidência.

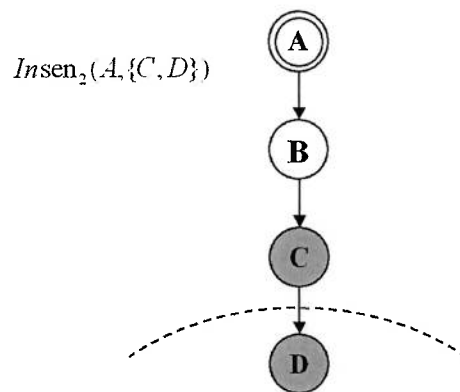


Figura 3.2. $Insen_2$.

Na rede acima, sendo A a probabilidade de interesse e C e D evidências, o conjunto de nós $Insen_2$ é formado pelos nós C e D . Ou seja, variação nos valores de probabilidades do nó D e de qualquer nó que viesse a existir como descendente de C não influenciaria o valor da probabilidade de interesse A , pois o nó descendente C é uma evidência.

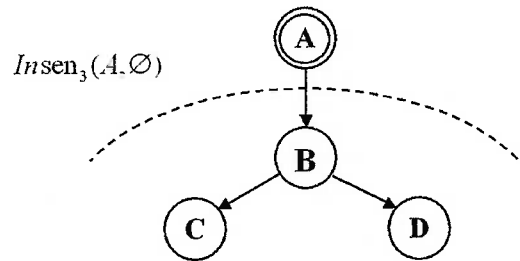


Figura 3.3. $Insen_3$.

Na rede acima, sendo A a probabilidade, o conjunto de nós $Insen_3$ é formado pelos nós B , C e D . Ou seja, variação nos valores de probabilidades de todos os descendentes do nó A não influenciaria o valor da probabilidade de interesse A , pois nenhum descendente é evidência.

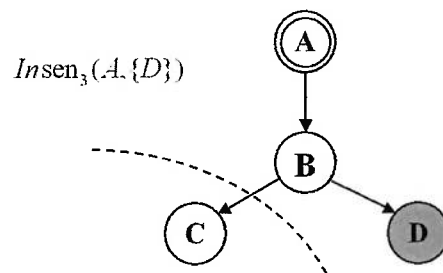


Figura 3.4. $Insen_3$.

Na rede acima, sendo A a probabilidade, o conjunto de nós $Insen_3$ é formado pelo nó C . Ou seja, variação nos valores de probabilidades de todos os descendentes que não sejam evidência ou não levem a uma evidência, não influenciaria o valor da probabilidade de interesse.

3.2.2 Relações de funções

O segundo princípio do método proposto por Coupé *et al* [7] é a observação de que a relação entre uma probabilidade de interesse e um parâmetro de uma rede Bayesiana pode ser expressa como um quociente de duas funções lineares.

Para esta seção as seguintes definições serão usadas:

1. Define-se x_{ik} como um valor de uma variável X_i e π_i uma combinação arbitrária de valores do conjunto de pais $pa(X_i)$.
2. Define-se também $y = p(a|E)$, onde a é o valor da variável de interesse A na derivada ($A \in \mathbf{X}$).

Assume-se que numa derivação, em se variando o parâmetro $x = p(x_{ik}|\pi)$ todas as outras probabilidades $p(x_{il}|\pi)$ são co-variadas de acordo pelo escalonamento da razão entre as massas de probabilidade restantes. Isto é necessário para se manter a normalização. Mais formalmente, seja o domínio da variável X_i indicado por $dom(X_i) = \{x_{i1}, \dots, x_{im}\}$, $m \geq 1$. Os parâmetros $p(x_{il}|\pi)$, $k \neq l$, são funções de x :

$$p(x_{il}|\pi)(x) = \begin{cases} x & \text{se } l = k \\ p(x_{il}|\pi)(x) \frac{1-x}{1-p(x_{ik}|\pi)} & \text{se } l \neq k \end{cases}$$

com $x = p(x_{ik}|\pi) < 1$.

Tem-se que [7, 21]:

Teorema 1 *Seja p a função de probabilidade definida por uma rede Bayesiana sobre um conjunto de variáveis \mathbf{X} . Seja $y = p(a|E)$ e $x = p(x_{ik}|\pi)$ como indicados anteriormente. Então,*

$$y = \frac{p(a, E)(x)}{p(E)(x)} = \frac{\alpha x + \beta}{\gamma x + \delta},$$

onde α , β , γ e δ são constantes com relação a x .

As constantes dessa função determinam diretamente as derivadas de interesse. Determinar essas constantes requer apenas três inferências na rede, que podem ser realizadas com qualquer algoritmo de inferência.

Exemplo 1 Usando a rede da Figura 3.5 como exemplo pode-se ilustrar o resultado do Teorema 1. Utiliza-se como probabilidade de interesse $p(a|c)$ e $p(c|-b)$ como probabilidade condicional em estudo. Então:

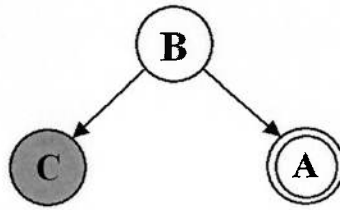


Figura 3.5. Exemplo de rede.

$$p(a|c) = \frac{p(a,c)}{p(c)}.$$

O numerador desta equação é igual a:

$$\begin{aligned} p(a,c) &= \sum_B p(c|B) \cdot p(a|B) \cdot p(B) \\ &= p(c|\neg b) \cdot p(a|\neg b) \cdot p(\neg b) + \\ &\quad + p(c|b) \cdot p(a|b) \cdot p(b) \\ &= \alpha \cdot x + \beta \end{aligned}$$

$$\text{onde } \alpha = p(a|\neg b) \cdot p(\neg b)$$

$$\text{e } \beta = p(c|b) \cdot p(a|b) \cdot p(b).$$

O denominador da equação é igual a:

$$\begin{aligned} p(c) &= \sum_{A,B} p(c|B) \cdot p(A|B) \cdot p(B) \\ &= \sum_B p(c|B) \cdot p(B) \\ &= p(c|\neg b) \cdot p(a|\neg b) + \\ &\quad + p(c|b) \cdot p(b) \\ &= \gamma \cdot x + \delta \\ &\text{onde } \gamma = p(\neg b) \\ &\text{e } \delta = p(c|b) \cdot p(b). \end{aligned}$$

E ainda mais, Coupe demonstra [7] que se a probabilidade condicional em estudo pertencer a um nó do conjunto de sensibilidade que não possui algum descendente observado, a função se reduz a uma função linear e não mais um quociente de duas funções.

Proposição 1 *Seja p a função de probabilidade definida por uma rede Bayesiana sobre um conjunto de variáveis X e seja $y = p(a|E)$ e $x = p(x_{ik}|\pi)$ como*

indicados anteriormente com a adição de que $\sigma(X_i) \cap O = \emptyset$ Então,

$$y = \frac{p(a, E)(x)}{p(E)(x)} = \alpha x + \beta,$$

onde α e β são constantes com relação a x .

Exemplo 2 Usando novamente a rede da Figura 3.5 como exemplo pode-se ilustrar o resultado da Proposição. Utiliza-se como probabilidade de interesse $p(a|c)$ e $p(a|b)$ como probabilidade condicional em estudo. Então:

$$p(a|c) = \frac{p(a,c)}{p(c)}.$$

O numerador da equação é igual a:

$$\begin{aligned} p(a, c) &= \\ &= \sum_B p(c|B) \cdot p(a|B) \cdot p(\mathbf{B}) \\ &= p(c|\neg b) \cdot p(a|\neg b) \cdot p(\neg b) + \\ &+ p(c|b) \cdot p(a|b) \cdot p(b) \\ &= \alpha' \cdot x + \beta' \end{aligned}$$

$$\text{onde } \alpha' = p(c|b) \cdot p(b)$$

$$\text{e } \beta' = p(c|\neg b) \cdot p(a|\neg b) \cdot p(\neg b).$$

O denominador da equação é igual a:

$$\begin{aligned} p(c) &= \\ &= \sum_{A,B} p(c|B) \cdot p(A|B) \cdot p(\mathbf{B}) \\ &= \sum_B p(c|B) \cdot p(B) \\ &= p(c|\neg b) \cdot p(a|\neg b) + \\ &+ p(c|b) \cdot p(b) \\ &= \gamma' \\ p(a|c) &= \frac{\alpha'x + \beta'}{\gamma'} = \alpha x + \beta \end{aligned}$$

Vale observar que, no caso especial em que nenhum nó da rede é observado, a segunda proposição implica que a probabilidade de interesse se relaciona linearmente com toda probabilidade condicional de cada nó do conjunto de sensibilidade em estudo.

Concluindo, este método argumenta que uma probabilidade de interesse de uma rede Bayesiana se relaciona com a probabilidade condicional em estudo por uma simples função matemática. O conhecimento dessa função permite uma considerável redução na carga computacional de uma derivação de uma rede Bayesiana já que apenas as constantes dessa função necessitam ser conhecidas. Essas constantes podem ser determinadas pela inferência da probabilidade de interesse da rede para um pequeno número de valores da probabilidade condicional em estudo e resolvendo o sistema de equações resultantes; variação sistemática do valor da probabilidade condicional não é mais necessário. Para uma probabilidade condicional que se relaciona linearmente com a probabilidade de interesse, duas inferências são suficientes; para todas as outras probabilidades condicionais, três inferências são necessárias. (Ou seja, com dois pontos definimos uma reta e com três definimos uma curva). Na pior das hipóteses serão necessárias três inferências na rede Bayesiana por probabilidade condicional em estudo.

Esquema da demonstração Mais uma vez usando a rede da Figura 3.5 como exemplo, utiliza-se como probabilidade de interesse $p(a|c)$ e $p(c|-b)$ como probabilidade condicional em estudo. Então:

$$p(a|c) = \frac{\alpha x + \beta}{\gamma x + \delta} = \frac{\alpha' x + \beta'}{x + \delta'}$$

Usando $\{x = 0, 2\}$, $\{x = 0, 4\}$ e $\{x = 0, 6\}$ para a probabilidade condicional em estudo e $p(c|b) = 0, 8$, $p(a|b) = 0, 096$, $p(a|-b) = 0, 905$ e $p(b) = 0, 01$ obtém-se:

$$p(a|c)_{x=0,2} = \frac{\alpha' \cdot 0,2 + \beta'}{0,2 + \delta'} = 0, 87356,$$

$$p(a|c)_{x=0,4} = \frac{\alpha' \cdot 0,4 + \beta'}{0,4 + \delta'} = 0, 88897 \text{ e}$$

$$p(a|c)_{x=0,6} = \frac{\alpha' \cdot 0,6 + \beta'}{0,6 + \delta'} = 0, 89424.$$

A partir desses valores obtemos três equações lineares:

$$\begin{cases} 0, 2\alpha' + \beta' - 0, 87356\delta' - 0, 2 \cdot 0, 87356 = 0 \\ 0, 4\alpha' + \beta' - 0, 88897\delta' - 0, 4 \cdot 0, 88897 = 0 \\ 0, 6\alpha' + \beta' - 0, 89424\delta' - 0, 6 \cdot 0, 89424 = 0 \end{cases}$$

Resolvendo o sistema linear obtemos: $\alpha' = 0, 905$, $\beta' = 0, 00061$ e $\delta' = 0, 00789$.

3.3 Método proposto por Kjaerulff e van der Gaag

O método revisto na seção anterior tem uma complexidade computacional que é consideravelmente menor do que a abordagem de força bruta. Força bruta sistematicamente varia os parâmetros, já o método do quociente de duas funções lineares realiza três inferências na rede para cada variável em estudo. Entretanto esse método pode ainda consumir bastante tempo: para uma rede de tamanho grande o número de variáveis em estudo chega facilmente a casa das centenas. Nessa situação o método requer várias centenas de inferências na rede podendo consumir tanto tempo que se torna impraticável.

Um método mais eficiente proposto por Kjaerulff e van der Gaag [21] é um aprimoramento do método proposto por Coupé *et al* [7]. Ele também utiliza o conceito de conjunto de sensibilidade demonstrado anteriormente e também o fato da probabilidade de interesse se relacionar como um quociente de duas funções em relação a probabilidade condicional em estudo. A diferença é que este método foi desenvolvido para redes Bayesianas na sua representação em árvore de junção. A vantagem deste método para o cálculo de derivadas é, como será demonstrado a seguir, o fato dele requerer apenas três propagações na árvore de junção. Com três propagações o método determina os coeficientes de todas as possíveis funções que determinam o efeito das probabilidades condicionais na probabilidade de interesse.

O método é baseado na idéia de que, numa árvore de junção, a expressão para $p(a, E)$ e $p(E)$ em função de x pode ser obtida do potencial de um clique contendo tanto a variável x como seus pais. O seguinte teorema apresentado por Kjaerulff e van der Gaag [21] detalha como os coeficientes podem ser calculados.

Teorema 2 *Seja p a função de probabilidade definida por uma rede Bayesiana e seja T a representação da rede na sua forma de árvore de junção. Seja $y = p(a|E)$ e $x = p(x_{ik}|\pi)$ como indicados anteriormente. Suponha que, em T , uma propagação partindo da extremidade tenha sido realizada em direção ao clique que contém a variável de interesse A ; suponha que subsequente uma propagação em direção à extremidade tenha sido realizada desse clique com o valor a para A . Seja Q um clique em T contendo tanto a variável X_i e seus pais $pa(X_i)$; seja $\phi_Q = p(Q, a, E)$ o potencial do clique Q depois das*

propagações citadas acima. Então, $p(a|E)(x) = \alpha x + \beta$ onde:

$$\alpha = \frac{\sum_{Q:x_{ik},\pi} \phi_Q}{p(x_{ik}|\pi)} - \sum_{l \neq k} \frac{\sum_{Q:x_{il},\pi} \phi_Q}{1 - p(x_{ik}|\pi)} \quad (3.1)$$

$$\beta = \sum_{l \neq k} \frac{\sum_{Q:x_{il},\pi} \phi_Q}{1 - p(x_{ik}|\pi)} + \sum_{Q:pa(X_i) \neq \pi} \phi_Q \quad (3.2)$$

Suponha agora que a evidência E tenha sido processada em T por uma propagação do topo em direção a raiz e subsequente propagação da raiz em direção ao topo. Seja $\phi_Q^* = p(Q, E)$ o potencial do clique Q depois das propagações citadas acima. Então, $p(E)(x) = \gamma x + \delta$ onde:

$$\gamma = \frac{\sum_{Q:x_{ik},\pi} \phi_Q^*}{p(x_{ik}|\pi)} - \sum_{l \neq k} \frac{\sum_{Q:x_{il},\pi} \phi_Q^*}{1 - p(x_{ik}|\pi)} \quad (3.3)$$

$$\delta = \sum_{l \neq k} \frac{\sum_{Q:x_{il},\pi} \phi_Q^*}{1 - p(x_{ik}|\pi)} + \sum_{Q:pa(X_i) \neq \pi} \phi_Q^* \quad (3.4)$$

O teorema acima fornece a base para o método de computar os coeficientes das funções que expressam a probabilidade de interesse $y = p(a|E)$ em termos de todos os possíveis parâmetros x . O método é composto dos seguintes passos [21]:

Algoritmo do método proposto por Kjaerulff e van der Gaag

1. Insira a evidência e na árvore de junção realize uma propagação em direção à raiz e outra no sentido oposto.
 2. Compute os coeficientes γ e δ usando as Equações (3.3) e (3.4) do Teorema 2, para todos os parâmetros relevantes, localmente por clique.
 3. Realize uma terceira propagação desta vez no sentido oposto da raiz a partir de um clique que contenha a variável de interesse A , com evidência adicional $\{A = a\}$.
 4. Compute os coeficientes α e β usando as Equações 3.1 e 3.2 do teorema 2, para todos os parâmetros relevantes, localmente por clique.
-

Este último método requer apenas três propagações na rede (na sua representação *árvore de junção*) para estabelecer todas as derivadas para uma probabilidade posterior marginal específica.

Através das duas primeiras propagações do algoritmo do método proposto por Kjaerulff e van der Gaag (passo número 1) se determina $p(X_i, E)$ para qualquer variável X_i . A terceira propagação é necessária para se determinar $p(a, X_i, E)$ para qualquer variável X_i . Com essas probabilidades e aplicando-se o Teorema 1 obtém-se a derivada da rede Bayesiana dada uma probabilidade de interesse para todas as possíveis probabilidades em estudo.

3.4 Comparação dos métodos exatos

Para dar algumas impressões quanto aos “custos” computacionais de derivadas em rede Bayesianas vamos levantar três exemplos reais:

- Alarm-network para monitoramento de paciente contém 37 nós, para os quais 571 valores de probabilidades foram especificados [2].
- A Wilson-network para diagnosticar a doença Wilson, com 21 nós e 162 valores [22].
- VSD-network que fornece prognóstico para crianças com *ventricular septum defect*, que contém 38 nós e 807 valores de probabilidades [24].

Utilizando-se variação sistemática, se para cada valor for realizada uma variação de 0 a 1 em passos de 0,1 obtém-se respectivamente 6281, 1782 e 8877 inferências nas redes.

Desconsidere agora as probabilidades que estão d-separadas da probabilidade de interesse, ou seja, levando-se em conta apenas o conjunto de sensibilidade para a probabilidade de interesse (na rede Alarm a probabilidade de interesse é um entre os oito diagnóstico disponíveis na rede, na rede Wilson é o diagnóstico da doença Wilson e na rede VSD é o prognóstico para crianças de defeito do septo ventricular. Se nenhuma observação foi determinada as redes Alarm, Wilson e VSD possuem apenas 3, 6, e 151 valores de probabilidades que influenciam a probabilidade de interesse respectivamente. Com observações de um paciente típico, o número de valores de probabilidades que influenciam aumentam respectivamente para 54, 32 e 491. Vale ressaltar que nas três redes de exemplo inserindo-se evidências o conjunto

Tabela 3.1. Tabela comparando os métodos de derivadas exatos.

Rede (número total de probabilidades)	Número de observações	Número de valores que influenciam	Número de inferências			
			Varição sistemática	Conjunto de sensibilidade	Quociente de duas funções lineares	Árvore de junção
ALARM (571)	0	3	6281	33	6	3
	6	54	6281	594	162	3
WILSON (162)	0	6	1782	66	12	3
	6	32	1782	352	96	3
VSD (807)	0	151	8877	1661	302	3
	9	491	8877	5401	1473	3

de sensibilidade aumenta. Isso se deve ao fato de que as evidências estão diminuindo o $Insen_3(V_r, O)$ (definido anteriormente).

Utilizando-se variação sistemática apenas no conjunto de sensibilidade, obtém-se respectivamente 33, 66 e 1661 inferências nas redes se não houver nenhuma observação e respectivamente 594, 352 e 5401 inferências nas redes com as observações de um paciente típico.

Utilizando-se agora todo o método do quociente de duas funções lineares, ou seja, determinação do conjunto de sensibilidade e que a relação entre uma probabilidade de interesse e um parâmetro de uma rede Bayesiana pode ser expressa como um quociente de duas funções lineares. Obtém-se respectivamente 6, 12 e 302 inferências nas redes se não houver nenhuma observação e respectivamente 162, 96 e 1473 inferências nas redes com as observações de um paciente típico.

Finalmente utilizando-se o método proposto por Kjaerulff e van der Gaag são necessárias apenas três inferências na árvore de junção (que é construída levando-se em conta o conjunto de sensibilidade e não toda a rede). Observa-se que o número de inferências necessárias pelo método proposto por Kjaerulff e van der Gaag é, para as três redes de exemplo, inferior a 0,2% do número de inferências que seriam necessárias através de uma variação sistemática nas redes.

A Tabela 3.1 (estendida a partir da tabela de comparação existente em [7]) resume os resultados obtidos para cada uma das três redes de exemplo.

3.5 Redes Bayesianas com abordagem diferencial

Os métodos de diferenciação revistos até agora utilizam inferência exata para o cálculo. Existem outros métodos exatos e também métodos aproximados.

Como referência Chan e Darwiche apresentam [5] uma abordagem para realizar inferência em redes Bayesianas que é baseada em derivadas parciais. De acordo com essa abordagem, uma rede Bayesiana é compilada em um polinômio de múltiplas variáveis e então são computadas as derivadas parciais desse polinômio com respeito a cada variável.

O novo método para diferenciação proposto neste trabalho é um método exato baseado em eliminação de variáveis. O método exato de Chan e Darwiche assim como métodos aproximados não são utilizados (nem partes desses métodos) e estão portanto fora do escopo deste trabalho.

3.6 Aplicações

Como já visto no Capítulo 2 uma rede Bayesiana é composta de duas partes: parte qualitativa formada pelo grafo direcionado acíclico; e parte quantitativa formada pelas distribuições de probabilidades condicionais associadas a cada um dos nós.

Com base na bibliografia consultada [7], construir a parte quantitativa de uma rede Bayesiana é a tarefa mais difícil e a que consome mais tempo. Esta tarefa significa atribuir valores para as diversas probabilidades condicionais de todas as variáveis representadas na rede e em redes reais o número de probabilidades envolvido é alto (como visto na Tabela 3.1).

Obter os valores das probabilidades não é uma tarefa simples. Apesar de existirem, em vários domínios, informações abundantes na literatura ou em dados estatísticos, acontece com frequência que essas informações não podem ser aplicadas diretamente na rede. Na literatura médica, por exemplo, se encontram probabilidades de sintomas dados uma doença, mas é raro encontrar a probabilidade de um sintoma ocorrer dado a ausência de uma doença e também probabilidades condicionais para estados intermediários da doença. A maioria das probabilidades da rede precisa então ser fornecida por especialistas (*domain experts*). Além de consumir muito tempo, existe a relutância dos especialistas em atribuir valores exatos para as

probabilidades, assim como a possibilidade de especialistas atribuírem valores diferentes para a mesma probabilidade. É comum que o valor de uma probabilidade ser um intervalo ao invés de um valor exato.

Diferenciação de parâmetros de redes Bayesianas é uma atividade que é de grande utilidade em pelo menos duas aplicações: *Análise de sensibilidade de redes Bayesianas* e *Otimização local de parâmetros*. Essas duas aplicações abordam, de formas distintas, a existência de intervalos possíveis para os valores das probabilidades.

Nas duas subseções seguintes serão vistos em mais detalhes essas duas aplicações.

3.6.1 Análise de sensibilidade de redes Bayesianas

Determinar com precisão o valor de uma probabilidade ou no caso de uma rede real centenas de valores, pode vir a ser muito custoso. Entretanto, no geral, nem toda probabilidade vai requerer o mesmo nível de precisão para se atingir um comportamento satisfatório da rede. Para se determinar o nível de precisão requerido para as diversas probabilidades condicionais da rede, pode-se realizar uma análise de sensibilidade [7, 5].

Numa análise de sensibilidade deseja-se descobrir qual é a influência que a variação de uma probabilidade em estudo (ou de um grupo de probabilidades) tem em uma probabilidade de interesse. Ou seja, deseja-se saber a derivada da rede: uma derivada onde a saída (ou valor da derivada) é o valor da probabilidade de interesse e se está derivando em relação à probabilidade em estudo. Algumas probabilidades vão demonstrar um efeito considerável, enquanto que outras quase não revelarão influência alguma. Para as variáveis que influenciam menos, os valores iniciais podem ser suficientes. Já para as variáveis que exercem maior influência, um refinamento pode ser benéfico. Dado o tempo limitado e custoso dos especialistas, a atenção pode ser focada nas probabilidades para as quais, através de uma análise de sensibilidade, o sistema mostrou ser mais sensível [7].

Após se determinar quais são os valores de probabilidade que mais influenciam a probabilidade de interesse, esses valores são refinados e conseqüentemente a sistema como um todo. Uma vez refinado o sistema pode mostrar novamente várias diferentes sensibilidades. Com base nessas observações é proposto um procedimento para o refinamento das probabilidades que está descrito na Figura 3.6 [7].

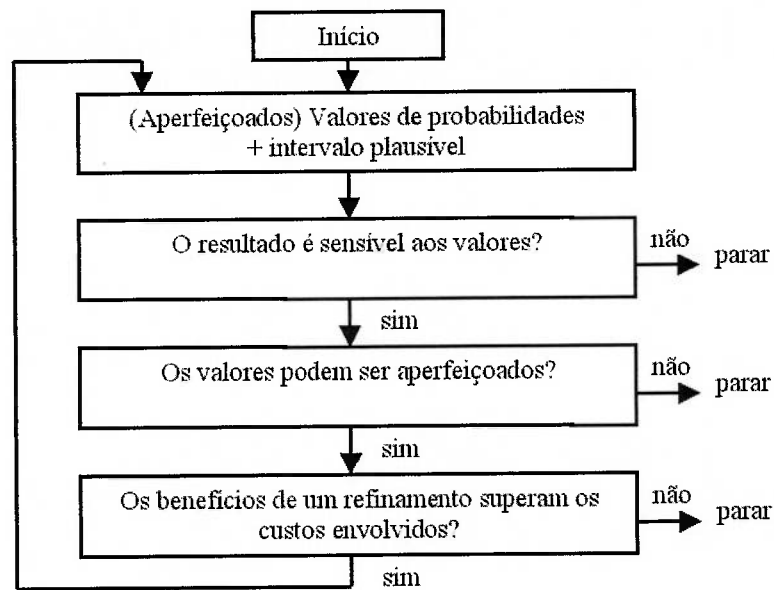


Figura 3.6. Procedimento para refinamento das probabilidades.

Esse procedimento é iniciado com a determinação inicial (e provavelmente com grande incerteza) de todas as probabilidades condicionais dos parâmetros da rede em construção. Começando com esses valores iniciais de probabilidades é realizada uma análise de sensibilidade na rede. Com base nessa análise os valores das probabilidades de maior influência são refinados. Interativamente é realizada análise de sensibilidade e refinamento das probabilidades até que seja atingido uma das seguintes situações: obtenção de comportamento satisfatório da rede; o custo de um refinamento das probabilidades supere os benefício de uma maior precisão; ou a impossibilidade de se obter um maior refinamento devido a falta de conhecimento.

3.6.2 Otimização local de parâmetros

Em muitas ocasiões o valor exato de uma probabilidade $p(X_i|pa(X_i))$ pode não ser conhecido. Isso ocorre quando o especialista construindo a rede tem informações vagas ou insuficientes; ou quando um conjunto de especialistas diverge sobre os parâmetros de uma rede; ou quando dados experimentais são coletados e não são suficientes para estimar um único valor de probabilidades. Como já falado anteriormente não se tem um valor exato de probabilidade, mas sim um valor dentro de um intervalo possível.

Nesses casos, uma solução possível é considerar intervalos de probabilidades para

os valores $p(X_i|pa(X_i))$, e realizar inferências a partir desses intervalos [3, 9, 14].

A realização de inferências requer então a otimização de valores de probabilidades com as restrições prescritas por intervalos. Em tais casos, um método de inferência importante é a otimização local por método do gradiente, a qual exige o cálculo de derivadas em cada uma de suas iterações. Os resultados apresentados nesse trabalho podem então ser utilizados.

Capítulo 4

UM NOVO MÉTODO BASEADO EM ELIMINAÇÃO DE VARIÁVEIS

O Capítulo 3 deste trabalho visou introduzir alguns dos atuais métodos para se realizar derivadas em rede Bayesianas. Mostrou-se com alguns números que realizar a análise de sensibilidade executando apenas variação sistemática é na prática, com redes de tamanho reais, computacionalmente inviável. No método do quociente de duas funções lineares, mostrou-se os benefícios de uma análise primeiro da parte qualitativa da rede, sem envolver probabilidades. Tal análise reduz o número de nós a serem computados em uma derivada apenas aos nós que afetam a probabilidade de interesse (o conjunto de sensibilidade). Esse método mostrou também que a probabilidade de interesse de uma rede Bayesiana pode ser expressa como um quociente de duas funções lineares de uma probabilidade condicional em estudo, reduzindo dessa forma a no máximo três inferências por probabilidade condicional em estudo. O método seguinte proposto por Kjaerulff e van der Gaag [21] utilizando redes Bayesianas em sua representação por árvore de junção foi uma clara evolução do método Fracional Linear, utilizando todos os conceitos introduzidos no método anterior como base.

O método de Kjaerulff e van der Gaag é entretanto baseado e construído em cima de redes Bayesianas na sua representação na forma de *árvore de junção*. Ou seja, utiliza complexos conceitos e métodos da teoria de grafos.

A proposta deste trabalho é a de se adaptar o método de derivação proposto por Kjaerulff e van der Gaag, ao algoritmo de inferência em redes Bayesianas por

eliminação de variáveis [10]. Este novo algoritmo não utiliza conceitos de grafos, tais como triangulação e cliques, evitando dessa forma a utilização de complexos métodos da teoria de grafos. Ao invés disso, este algoritmo se foca apenas em densidades de probabilidades.

4.1 Algoritmo proposto

Como adaptar um método que utiliza redes Bayesianas na sua representação de árvore de junção para um algoritmo que utiliza árvore de *buckets*? A resposta é simples: o algoritmo de generalização de eliminação de variáveis também pode ser interpretado utilizando-se ferramentas de teoria de grafos. O resultado básico aqui é que toda árvore de *buckets* é também uma árvore de cliques (os cliques da árvore de triangulação são induzidos pelo ordenamento das variáveis) e portanto pode ser visto como uma *árvore de junção* [10].

Tendo isso em mente, propõe-se o seguinte algoritmo que adapta a derivação do método por *árvore de junção* na estrutura de *buckets* de *eliminação de variáveis*. O método é bastante similar a uma inferência em árvore de *buckets*, mas contém alguns passos adicionais para cálculo de valores usados em derivadas:

Algoritmo Derivadas em Redes Bayesianas Usando Eliminação de Variáveis

1. Gere um ordenamento para as N variáveis que são necessárias para a inferência, que não observadas e não questionadas (a variável de interesse na derivação A deve ser a variável questionada X_q na eliminação de variáveis).
2. Coloque todas as densidades da rede em um conjunto de densidades.
3. Primeira propagação.

Para i de 1 a N :

- a. Crie uma estrutura de dados B_i , chamada *bucket*, contendo:

- A variável X_i , chamada variável do *bucket*;
- Todas as densidades que contenham a variável do *bucket*, chamado de densidade do *bucket*, mas que ainda não foram colocadas em nenhum *bucket*;

- b. Multiplique as densidades em B_i . Armazene a densidade resultante não normalizada em B_i ; a densidade é chamada de *cluster* de B_i .
 - c. Elimine (somando) X_i do *cluster* B_i . Armazene a densidade resultante não normalizada em B_i ; a densidade é chamada de *separador* de B_i .
 - d. Coloque o *separador* do *bucket* no conjunto contendo densidades.
4. No final do processo colete as densidades que contém a variável questionadas num *bucket* B_q . Multiplique as densidades em B_q (e normalize o resultado para obter $p(A|E)$ com as condições iniciais da rede).
5. Segunda propagação.
- Para i de N a 1:
- a. Normalize o *cluster* de B_i em relação à variável do *bucket*, obtendo-se a densidade normalizada $p(X_i|\mathbf{S}_i, E_i)$;
 - b. Peça ao filho de B_i que forneça $p(\mathbf{S}_i|E_i)$. Esta densidade de probabilidade substitui o *separador* de B_i ;
 - c. Multiplique $p(\mathbf{S}_i|E_i)$ e o *cluster* normalizado do *bucket* $p(X_i|\mathbf{S}_i, E_i)$, obtendo $p(X_i, \mathbf{S}_i|E_i)$. Esta densidade de probabilidade substitui o *cluster* de B_i .
6. Compute os coeficientes γ e δ , usando as Equações (3.3) e (3.4), para todos os parâmetros relevantes, localmente por clique/*bucket*.
7. Terceira propagação.
- Realize outra propagação saindo da raiz em direção ao topo do clique/*bucket* B_q , desta vez com a evidência adicional $\{A = a\}$.
8. Compute os coeficientes α e β , usando as Equações (3.1) e (3.2), para todos os parâmetros relevantes, localmente por clique/*bucket*.
-

4.2 Comparação do método proposto com métodos existentes

Dos métodos exatos estudados os dois primeiros (variação sistemática e quociente de duas funções lineares) não possuem uma eficiência equivalente ao novo método

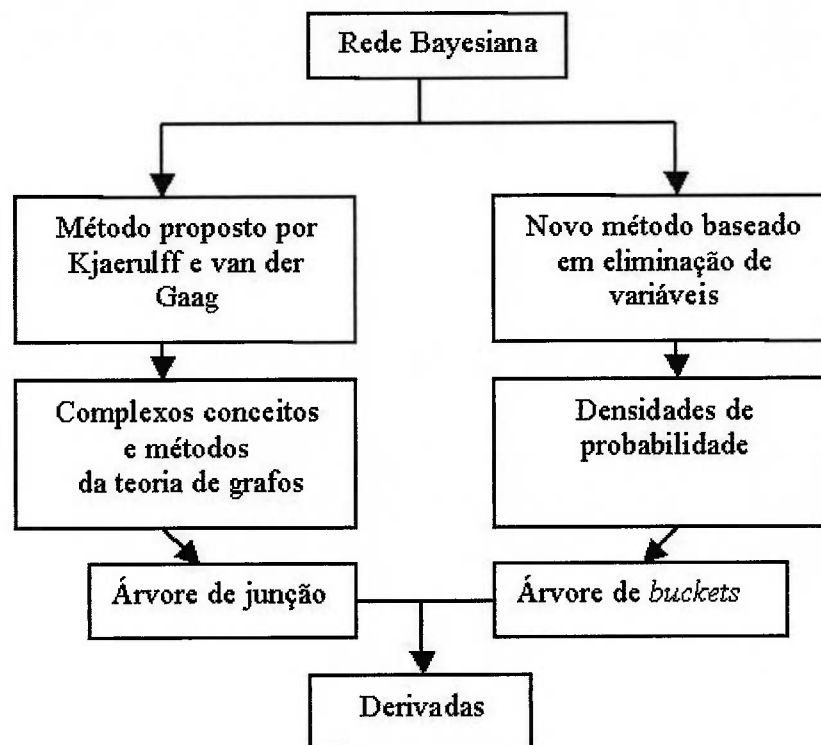


Figura 4.1. Comparação do novo método proposto com o método proposto por Kjaerulff e van der Gaag.

proposto. Como visto anteriormente o número de inferências necessárias nesses dois métodos, para uma rede de tamanho real, pode vir a ser centenas senão milhares de vezes maior do que no método proposto.

Dos métodos exatos estudados o que possui um desempenho que pode ser comparado é o método proposto por Kjaerulff e van der Gaag. Nesse método, assim como no novo método proposto são necessárias apenas três inferências para se determinar as derivadas de todos os possíveis probabilidades em estudo na probabilidade de interesse.

Uma vez que a árvore de junção do método proposto por Kjaerulff e van der Gaag e a árvore de clique do novo método baseado em eliminação de variáveis são equivalentes [10], comparar a eficiência desses dois métodos é na verdade comparar a eficiência da construção das árvores. Isso é verdade pois o algoritmo utilizado para se calcular a derivada (os quatro coeficientes das funções que expressam a probabilidade de interesse $y = p(a|E)$ em termos de todos os possíveis parâmetros x) é o mesmo nos dois métodos. As partes dos dois métodos cuja eficiência deve ser comparada estão ilustradas na Figura 4.1.

Dependendo da ordenação da rede inferida um método pode apresentar uma eficiência maior do que o outro. Como já citado anteriormente, no método de eliminação de variáveis, a ordenação das variáveis é arbitrária, mas diferentes ordenações levam a diferentes cargas computacionais. Infelizmente, a construção de uma ordenação ótima é conhecido como um problema NP-completo, então deve-se aceitar uma ordenação boa, não necessariamente ótima [10].

4.3 Demonstração do método

Nesta seção a teoria proposta é aplicada num exemplo simples. Para tanto é utilizada a rede da Figura 2.2 com os valores de probabilidades (considerando que todas as variáveis são booleanas) da Tabela 4.1.

Tabela 4.1. Tabela dos valores das probabilidades.

$p(a) =$	0,20	$p(d c, b) =$	0,80
		$p(d \neg c, b) =$	0,80
$p(c a) =$	0,20	$p(d c, \neg b) =$	0,80
$p(c \neg a) =$	0,05	$p(d \neg c, \neg b) =$	0,05
$p(b a) =$	0,80	$p(e c) =$	0,80
$p(b \neg a) =$	0,20	$p(e \neg c) =$	0,60

Também é assumido como evidência que a variável E é verdadeira. A probabilidade de interesse é $p(c|e)$. Vamos estudar a derivada em relação a três probabilidades: $p(e|\neg c)$, $p(c|\neg a)$ e $p(a)$.

Aplicando os conceitos de d-separação a rede é reduzida a três variáveis (A , C e E), gerando uma *árvore de buckets* com dois *buckets*, conforme Figura 4.2. A variável do *bucket* B_1 é A e a variável questionada C está no *bucket* B_q .

Utilizando o algoritmo de eliminação de variáveis inicia-se pelo *bucket* B_1 . A variável do *bucket* é A e as densidades do *bucket* são $p(A)$ e $p(C|A)$. O *cluster* de B_1 é obtido através da multiplicação das densidades e elimina-se (somando) a variável

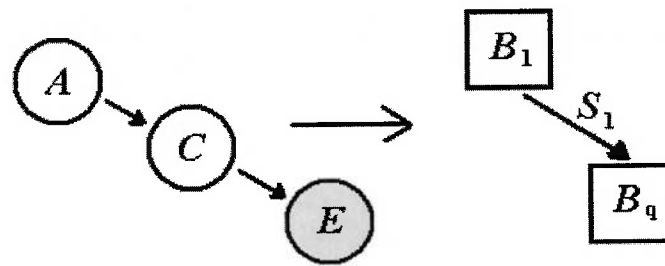


Figura 4.2. Transformando em árvore de *Buckets*.

A obtendo o separador S_1 .

$$\phi_1(A, C) \Rightarrow \begin{cases} \phi_1(a, c) = 0,04 \\ \phi_1(a, -c) = 0,16 \\ \phi_1(-a, c) = 0,04 \\ \phi_1(-a, -c) = 0,76 \end{cases}$$

$$S_1 = \sum_A \phi_1 \Rightarrow \begin{cases} S_1(c) = 0,08 \\ S_1(-c) = 0,92 \end{cases}$$

Passa-se agora para o *bucket* B_q . A variável do *bucket* é a variável questionada C e as densidades do *bucket* são $p(C)$ e o separador S_1 . Multiplicando as densidades obtém-se a variável questionada C .

$$\begin{aligned} \phi_q(c) &= 0,064 \\ \phi_q(-c) &= 0,552 \end{aligned} \text{ , normalizando obtemos}$$

$$\phi_q(c) = p(c|e) = 0,103896$$

$$\phi_q(-c) = p(-c|e) = 0,896104$$

Generalizando eliminação de variáveis é realizada a segunda propagação, desta vez partindo da raiz em direção ao topo e com o valor obtido para a variável questionada C .

Primeiro normaliza-se o *cluster* de B_1 em relação a variável A e depois multiplica-se o resultado por $p(C|E)$ obtido na primeira propagação que substitui o valor do

separador S_1 .

$$\phi_1(A, C) \Rightarrow \begin{cases} \phi_1(a, c) = 0,5 \\ \phi_1(a, \neg c) = 0,173913 \\ \phi_1(\neg a, c) = 0,5 \\ \phi_1(\neg a, \neg c) = 0,826087 \end{cases}$$

multiplicando por $p(C|E)$

$$\begin{cases} \phi_1(a, c) = 0,0519481 \\ \phi_1(a, \neg c) = 0,155844 \\ \phi_1(\neg a, c) = 0,0519481 \\ \phi_1(\neg a, \neg c) = 0,740260 \end{cases}$$

Pode-se obter o valor de $p(A|E)$ eliminando a variável C :

$$\sum_C \phi_1^* \Rightarrow \begin{cases} \phi_1^*(a) = 0,207792 \\ \phi_1^*(\neg a) = 0,792208 \end{cases}$$

Calcula-se os coeficientes γ e δ para as três probabilidades em estudo:

$$\begin{aligned} p(e|\neg c) &\Rightarrow \begin{cases} \gamma = 0,92 \\ \delta = 0,064 \end{cases} \\ p(c|\neg a) &\Rightarrow \begin{cases} \gamma = 0,259740 \\ \delta = 0,987013 \end{cases} \\ p(a) &\Rightarrow \begin{cases} \gamma = 0,0487013 \\ \delta = 1,99026 \end{cases} \end{aligned}$$

Realizando uma última propagação de B_q para B_1 com a evidência adicional $\{C = c\}$.

$$\begin{aligned} \phi_q &\Rightarrow \begin{cases} \phi_q(c) = 0,064 \\ \phi_q(\neg c) = 0 \end{cases} \\ \phi_1 &\Rightarrow \begin{cases} \phi_1(a, c) = 0,0519481 \\ \phi_1(a, \neg c) = 0 \\ \phi_1(\neg a, c) = 0,0519481 \\ \phi_1(\neg a, \neg c) = 0 \end{cases} \end{aligned}$$

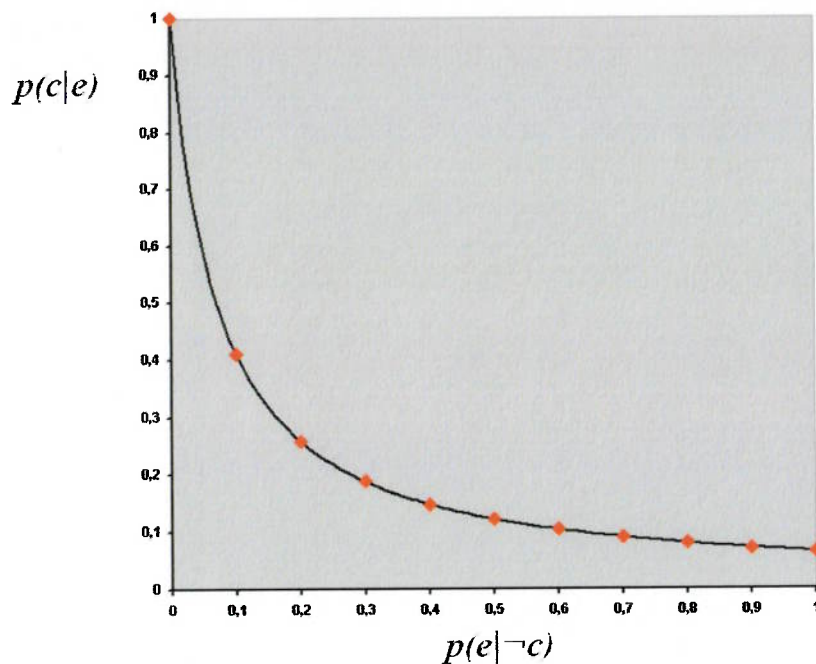


Figura 4.3. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(e|\neg c)$.

Percebe-se claramente nesta curva que a influência da variação do valor da probabilidade em estudo no valor da probabilidade de interesse é bastante acentuada e não linear.

Finalmente, computa-se os coeficientes α e β para as três probabilidades em estudo:

$$p(e|\neg c) \Rightarrow \begin{cases} \alpha = 0 \\ \beta = 0,064 \end{cases}$$

$$p(c|\neg a) \Rightarrow \begin{cases} \alpha = 1,03896 \\ \beta = 0,0519481 \end{cases}$$

$$p(a) \Rightarrow \begin{cases} \alpha = 0,194805 \\ \beta = 0,168831 \end{cases}$$

Para ilustrar a eficiência do método aqui utilizado foi realizada uma comparação com uma derivação através de força bruta (para tanto foi utilizado o programa JavaBayes). Para visualizar os resultados e a comparação realizada foram geradas as três curvas (utilizando-se os coeficientes obtidos) e em cada curva foram marcados os dez pontos da análise por força bruta. As Figuras 4.3, 4.4 e 4.5 trazem esses resultados para cada uma das variáveis em estudo.

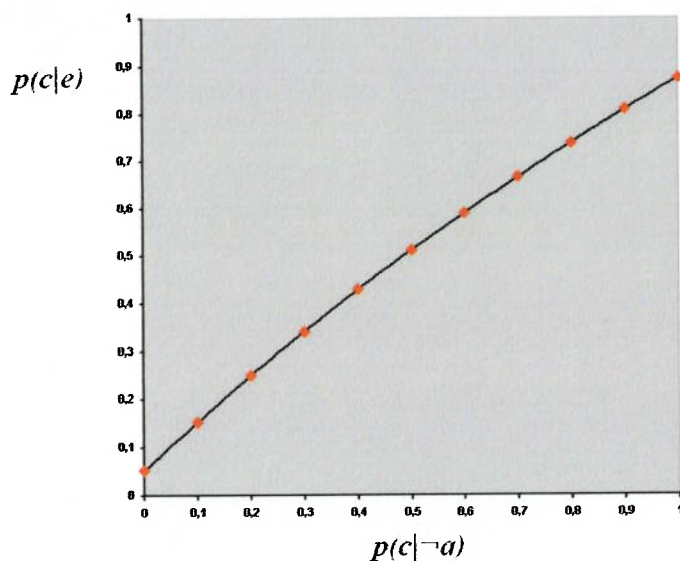


Figura 4.4. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(c|-a)$.

A influência da variação do valor deste probabilidade em estudo no valor da probabilidade de interesse é também bastante acentuada.

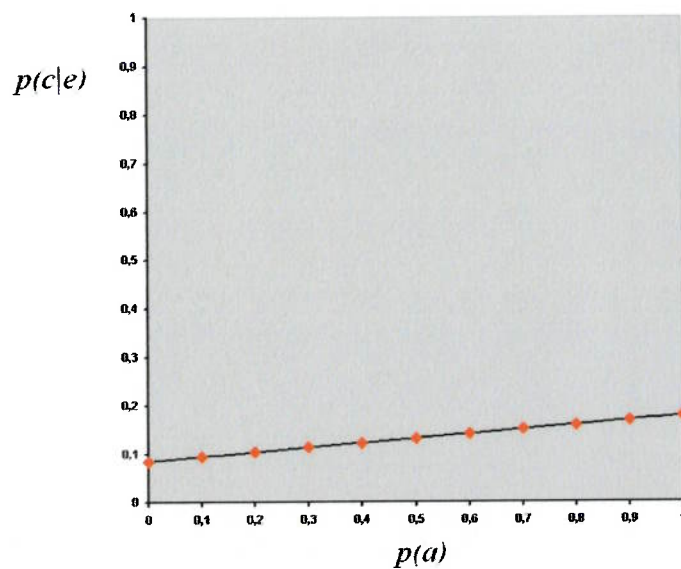


Figura 4.5. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(a)$.

Nesta última curva a influência da variação do valor da probabilidade em estudo é bem mais restrita do que nos dois casos anteriores.

4.4 Conclusões

Este novo método proposto para diferenciação de redes Bayesianas atinge os objetivos propostos:

- É um algoritmo de simples entendimento. O algoritmo utiliza apenas densidades de probabilidade ao invés de teoria de grafos (como é o caso do método de Kjaerulff e van der Gaag).
- É também de simples implementação. Este novo método foi implementado em Java como descrito no Capítulo 5.1.

Por ser um método exato de diferenciação ele possui a desvantagem de todo algoritmo exato de inferência: pode se tornar muito lento para redes de grande porte. No caso deste método, por ser baseado no algoritmo que generaliza eliminação de variáveis, a complexidade de tempo e de espaço pode ser exponencial no tamanho do maior clique existente no grafo da rede triangulado [12].

Capítulo 5

RESULTADOS EXPERIMENTAIS

O objetivo deste capítulo é mostrar cálculos usando o novo método proposto para o cálculo de derivadas em redes Bayesianas.

A primeira seção deste capítulo é sobre o programa desenvolvido em Java que implementa o novo método. Nas três seções seguintes estão indicados os resultados dos testes realizados com redes Bayesianas escolhidas utilizando-se o programa desenvolvido. A primeira rede escolhida é a rede de exemplo utilizada nos capítulos anteriores desta dissertação. As outras duas são redes reais com aplicações na medicina:

- Rede Alarm (37 nós). Rede real e em uso para monitorar pacientes na UTI.
- Rede Cardiomiopatia (28 nós). Rede real cujo problema médico representado é o de insuficiência cardíaca (cardiomiopatia).

Para a primeira rede escolhida com exemplo serão indicadas as derivadas em relação a todas as probabilidades em estudo. Isso é viável devido ao tamanho reduzido da rede. Já para as duas outras redes escolhidas, devido a seus tamanhos, serão apenas indicados alguns dos resultados obtidos, suficientes para demonstrar o funcionamento do novo método proposto e do programa desenvolvido.

Vale ressaltar também que para a primeira rede de exemplo todos os resultados foram validados através de cálculos manuais (Seção 4.3) e também pelo cálculo da derivada através do método de variação sistemática.

Já para as outras duas redes utilizadas, devido ao tamanho, comprovar os resultados obtidos através de cálculos manuais seria inviável. Validação completa pelo

método de variação sistemática também consumiria muito tempo. A validação foi então realizada apenas nos resultados obtidos para algumas das probabilidades em estudo (incluindo todas as que são mostradas nas seções seguintes). Para validar esses resultados obtidos utilizou-se o segundo princípio do método proposto por Coupé *et al* [7] e foram então inferidos apenas mais dois pontos das curvas obtidas (o primeiro era a inferência na condição inicial da rede).

5.1 Programa desenvolvido em Java

Além de apresentar um método exato para se obter derivadas em redes Bayesianas baseado no algoritmo de eliminação de variáveis, também foi um construído um programa, livremente distribuído, que tem como objetivo a implementação desse novo método.

Como citado na introdução, um dos objetivos deste trabalho é a implementação desse novo método e a integração ao sistema JavaBayes. JavaBayes é um pacote em Java livremente distribuído para a construção e manipulação de redes Bayesianas desenvolvido sob a orientação do pesquisador Fábio G. Cozman no Laboratório de Tomada de Decisão da Escola Politécnica.

O novo método foi implementado em Java e integrado ao pacote, aumentando sua capacidade. Dentre suas atividades de manipulação de redes Bayesianas adicionou-se a capacidade de derivar as redes Bayesianas.

Porém, mais do que integrar o novo método ao pacote JavaBayes este novo método foi desenvolvido como uma extensão à biblioteca EmBayes. EmBayes é a biblioteca desenvolvida na linguagem de programação Java para criar, manipular, avaliar e aprender redes Bayesianas que é utilizada pelo pacote JavaBayes.

Ser desenvolvido como uma extensão da biblioteca EmBayes possui a vantagem de permitir sua utilização em outros pacotes além do JavaBayes, como por exemplo no ambiente R ¹.

Além disso (assim como o pacote JavaBayes) essa biblioteca tem a características o código fonte aberto e a disponibilidade pela Internet.

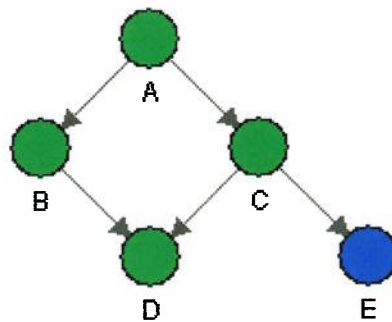
¹R é um ambiente usado para computação estatística e gráfica e está disponível em <http://www.r-project.org>.

Tabela 5.1. Tabela de coeficientes da rede de exemplo.

Variável em estudo	α	β	γ	δ
A=true	0,194805	0,064935	0,048701	0,99026
A=false	-0,194805	0,25974	-0,048701	1,038961
C=true A=true	0,25974	0,051948	0,064935	0,987013
C=true A=false	1,038961	0,051948	0,25974	0,987013
C=false A=true	-0,25974	0,311688	-0,064935	1,051948
C=false A=false	-1,038961	1,090909	-0,25974	1,246753
E=true C=true	0,12987	0	0,12987	0,896104
E=true C=false	0	0,103896	1,493506	0,103896

5.2 Rede de exemplo (5 nós)

A escolha da primeira rede Bayesiana para a obtenção de resultados experimentais não foi devido a sua complexidade, já que possui apenas cinco nós, mas por ter sido a rede utilizada ao longo desta dissertação como exemplo e para demonstrar o novo método.

**Figura 5.1.** Rede de exemplo.

Novamente é assumido como evidência que a variável E é verdadeira (nó em mais escuro na Figura 5.1) e que a probabilidade de interesse é $p(c|e)$. Vamos estudar a derivada em relação não apenas em relação às probabilidades $p(e|\neg c)$, $p(c|\neg a)$ e $p(a)$ como realizado na Seção 4.3, mas sim em relação a todas as probabilidades do conjunto de sensibilidade.

Os valores dos coeficientes obtidos para as probabilidades em estudo $p(e|\neg c)$,

$p(c|\neg a)$ e $p(a)$ foram os mesmos da Seção 4.3. Por isso gráficos dessas derivadas não serão aqui repetidos.

A Tabela 5.1 indica os valores dos coeficientes α , β , γ e δ para todas as possíveis variáveis em estudo.

As Figuras 5.2, 5.3, 5.4, 5.5 e 5.6 ilustram a derivada para as demais possíveis probabilidades em estudo.

5.3 Rede Alarm (37 nós)

A rede Alarm foi escolhida por ser uma rede real para monitorar pacientes na UTI.

A definição de todos os nós, variáveis e probabilidades pode ser encontrado no Apêndice A.2 e mais detalhes sobre a rede podem ser encontrados em [2].

Para este exemplo é assumido como evidência que a variável BP tem o valor Low (nó mais escuro na Figura 5.7) e que a probabilidade de interesse é $p(CO = Low|BP = Low)$.

Aplicando o conceito de d-separação a rede é reduzida a vinte e quatro variáveis. O conjunto de sensibilidade pode ser visto na Figura 5.8.

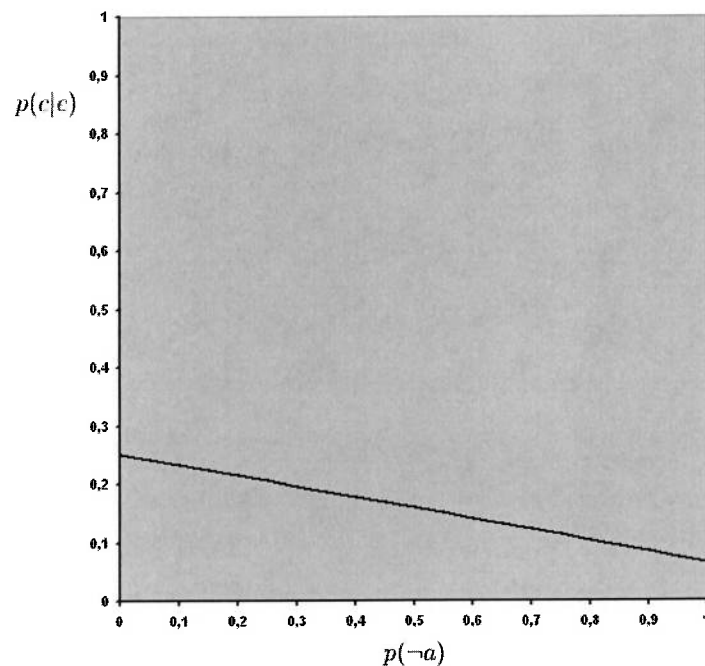


Figura 5.2. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(\neg a)$.

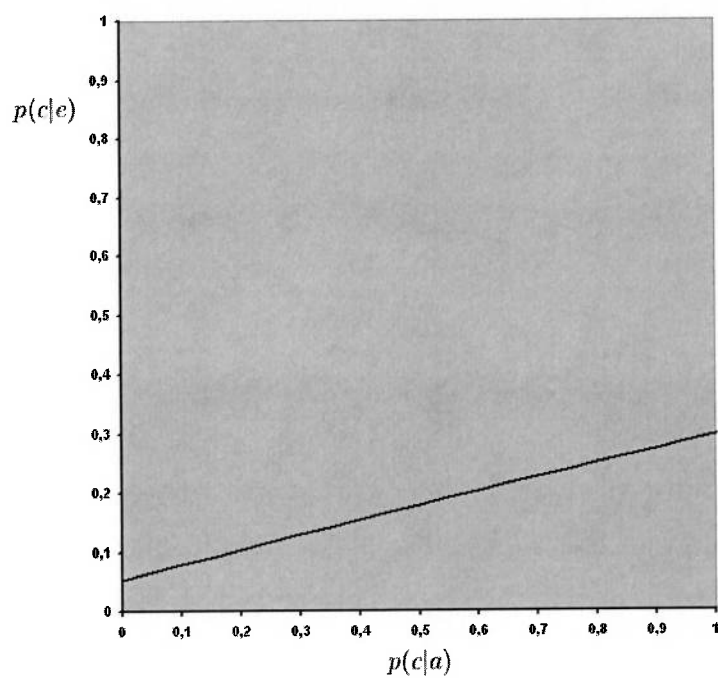


Figura 5.3. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(c|a)$.

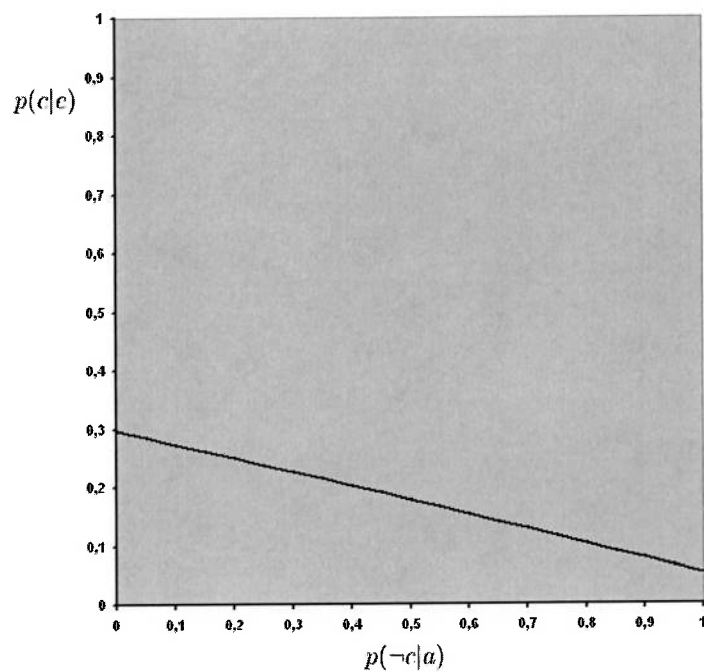


Figura 5.4. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(-c|a)$.

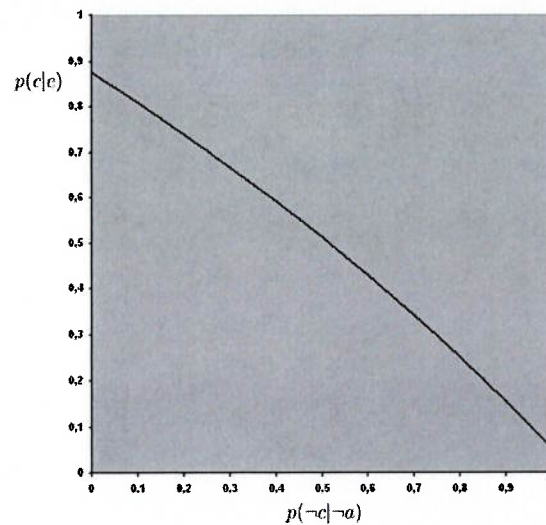


Figura 5.5. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(\neg c|\neg a)$.

Através de uma simples análise das figuras constata-se que os valores das probabilidades em estudo ($p(\neg c|\neg a)$ da figura acima, $p(e|\neg c)$ da Figura 4.3 e $p(c|\neg a)$ da Figura 4.4) são os que mais influenciam a probabilidade de interesse $p(c|e)$, sem levar em conta intervalos possíveis para os valores das probabilidades. A princípio esses três valores de probabilidades seriam o foco de trabalho de um possível refinamento da rede.

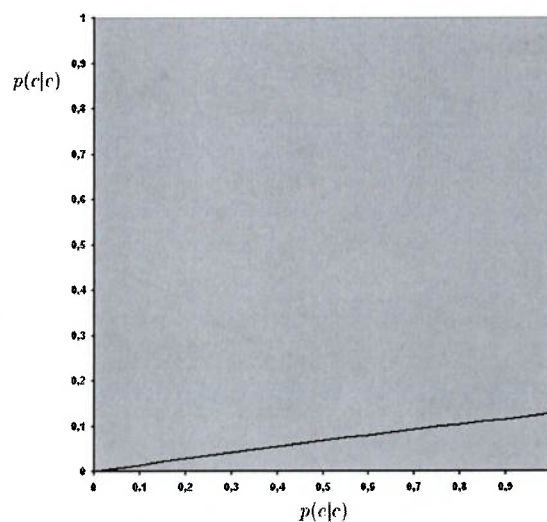


Figura 5.6. Curva de derivada: probabilidade de interesse $p(c|e)$ e probabilidade em estudo $p(e|c)$.

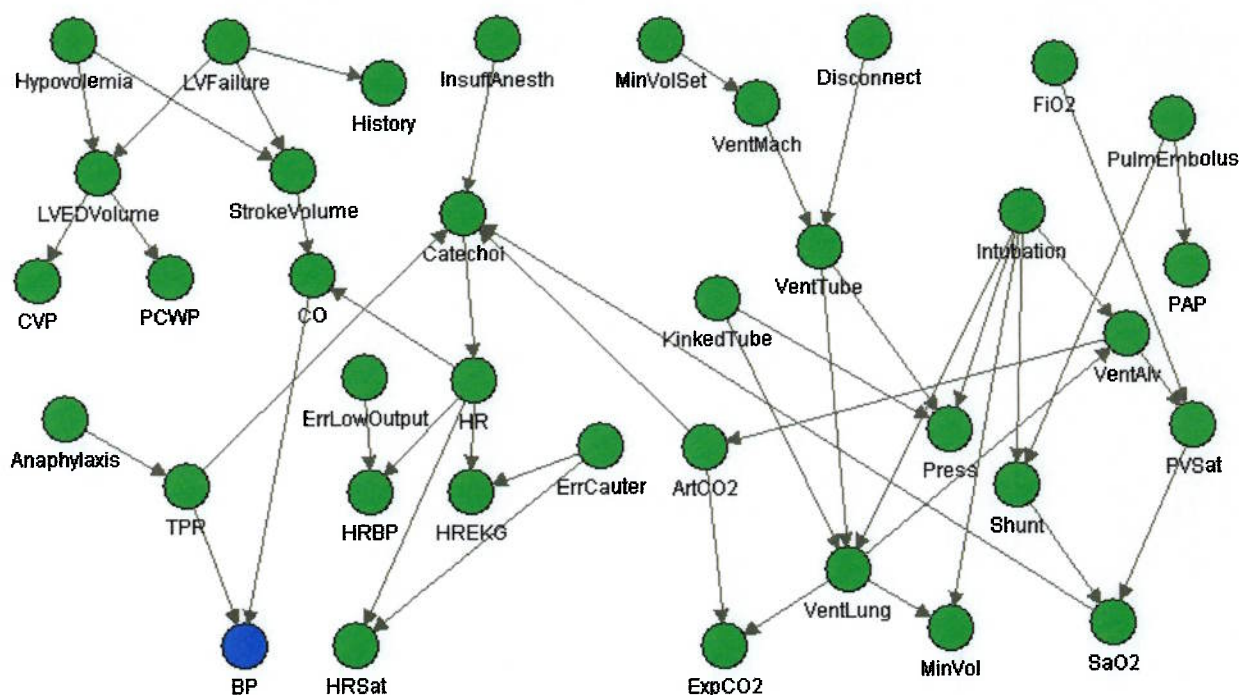


Figura 5.7. Rede Alarm.

Tabela 5.2. Tabela de coeficientes das três probabilidades em estudo.

Variável em estudo	α	β	γ	δ
Hypovolemia=True	1,227161	0,206473	0,733193	0,853361
StrokeVolume=Low LVFailure=False, Hypovolemia=False	1,061524	0,398829	0,634537	0,968273
Catechol=Normal InsuffA- nesth=False, SaO2=Low, TPR=Low, ArtCO2=Normal	0,003637	0,451724	0,00161	0,99992

Ao se aplicar o método proposto a árvore de junção é formada por 23 *buckets*. Como existem várias centenas de probabilidades em estudo serão indicados aqui apenas as duas probabilidades em estudo cuja variação mais influenciam a probabilidade de interesse (Figuras 5.9 e 5.10) e também uma probabilidade cuja variação quase não tem influência (Figura 5.11).

A Tabela 5.2 indica os valores dos coeficientes α , β , γ e δ para as três variáveis em estudo.

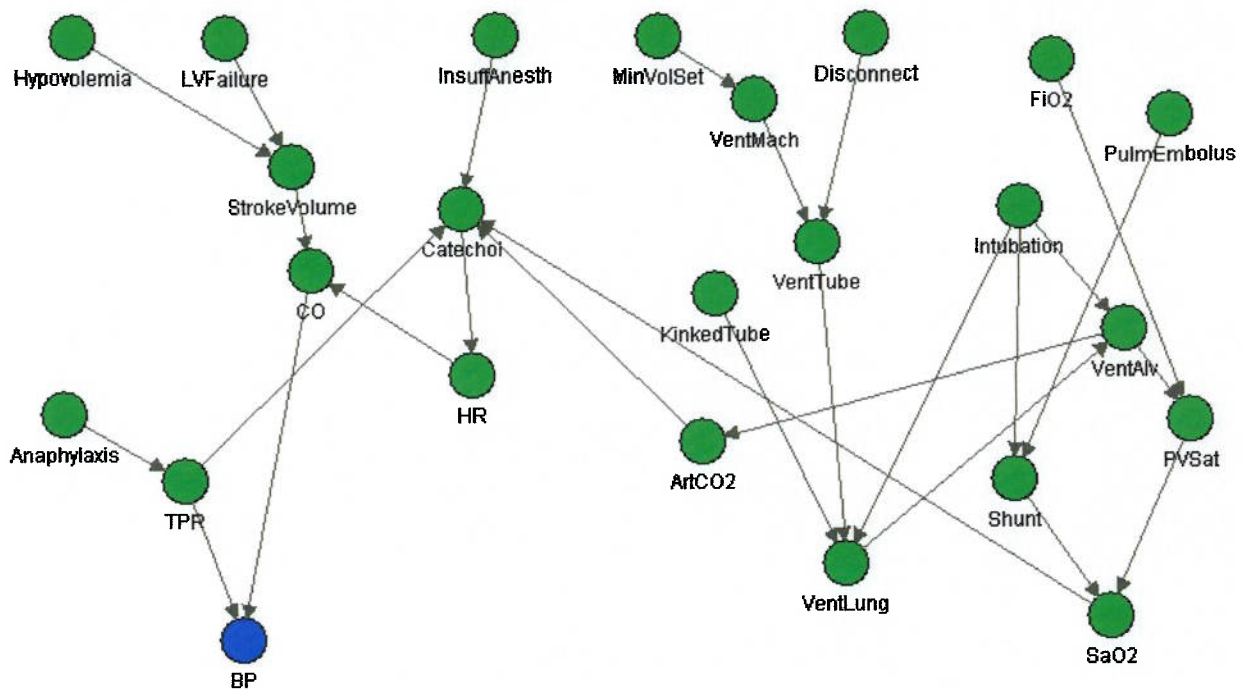


Figura 5.8. Rede Alarm - conjunto de sensibilidade.

Com $\{CO = Low\}$ como variável de interesse e $\{BP = Low\}$ como única evidência e aplicando o conceito de d-separação, temos uma redução das 37 variáveis iniciais para apenas 24 pertencentes ao conjunto de sensibilidade.

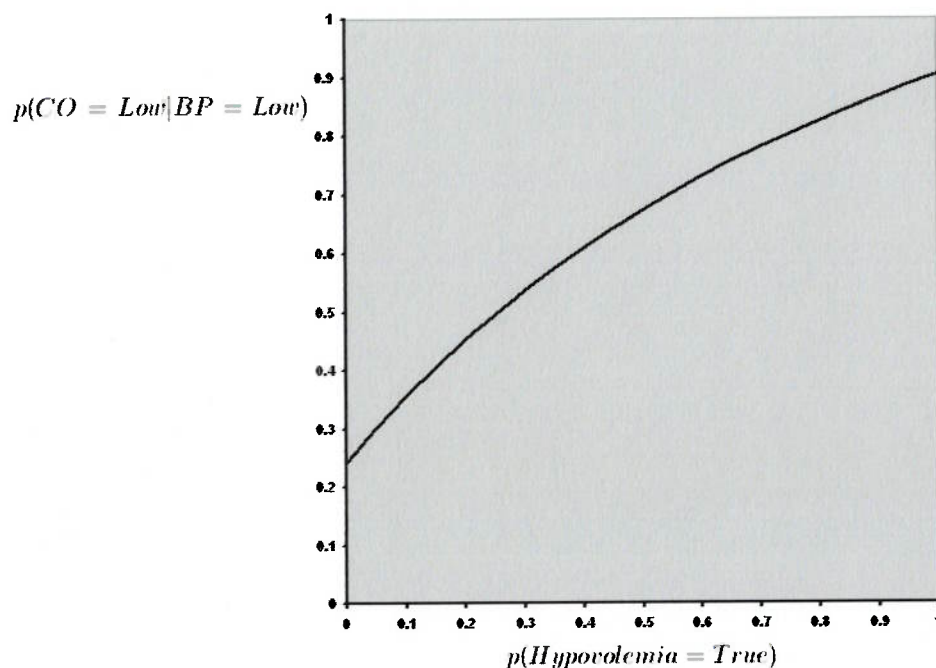


Figura 5.9. Curva de derivada: probabilidade de interesse $p(CO = Low | BP = Low)$ e probabilidade em estudo $p(Hypovolemia = True)$.

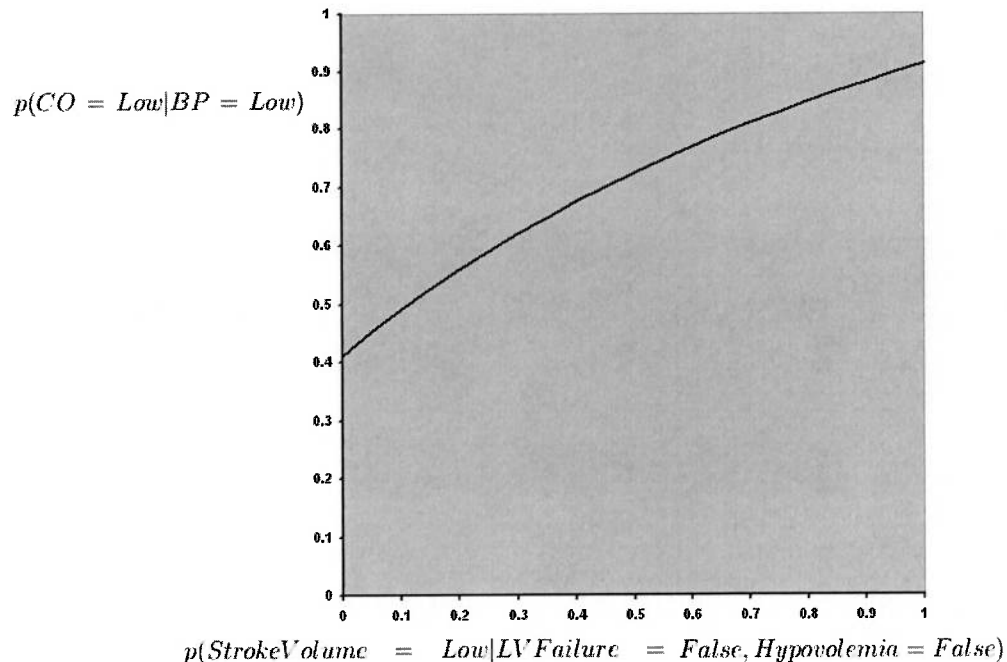


Figura 5.10. Curva de derivada: probabilidade de interesse $p(\text{CO} = \text{Low} | \text{BP} = \text{Low})$ e probabilidade em estudo $p(\text{StrokeVolume} = \text{Low} | \text{LV Failure} = \text{False}, \text{Hypovolemia} = \text{False})$.

As duas primeiras derivadas representadas na figura acima e na Figura 5.9 indicam que os valores das probabilidades em estudo $p(\text{Hypovolemia} = \text{True})$ e $p(\text{StrokeVolume} = \text{Low} | \text{LV Failure} = \text{False}, \text{Hypovolemia} = \text{False})$ possuem uma grande influência sobre a probabilidade de interesse $p(\text{CO} = \text{Low} | \text{BP} = \text{Low})$. Percebe-se também claramente que a influência é não linear. Já na Figura 5.11 a variação do valor da probabilidade em estudo $p(\text{Catechol} = \text{High} | \text{Insuff Anesth} = \text{False}, \text{SaO2} = \text{Low}, \text{TPR} = \text{Low}, \text{ArtCO2} = \text{Normal})$ praticamente não possui influência alguma no valor da probabilidade de interesse $p(\text{CO} = \text{Low} | \text{BP} = \text{Low})$. Num possível refinamento da rede Alarm os valores das duas primeiras probabilidades em estudo seriam o foco. Já o valor inicial da terceira probabilidade em estudo seria o suficiente.

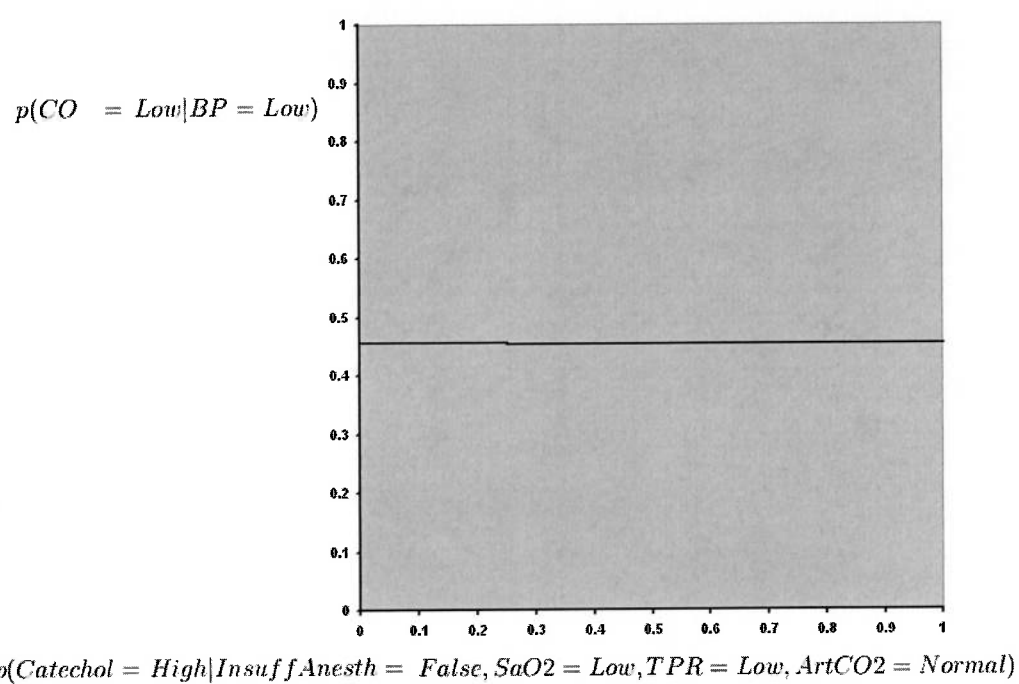


Figura 5.11. Curva de derivada: probabilidade de interesse $p(CO = Low | BP = Low)$ e probabilidade em estudo $p(Catechol = High | InsuffAnesth = False, SaO2 = Low, TPR = Low, ArtCO2 = Normal)$.

5.4 Rede Cardiomiopatia (28 nós)

Esta rede foi desenvolvida pelo colega André Saheki do Laboratório de Tomada de Decisão da POLI, em parceria com o Dr. Rodolfo Sharovsky do Hospital Universitário (HU-USP). O problema médico representado é o diagnóstico de insuficiência cardíaca (cardiomiopatia).

A definição de todos os nós, variáveis e probabilidades pode ser encontrado no Apêndice A.3 e mais detalhes sobre a rede podem ser encontrados em [27].

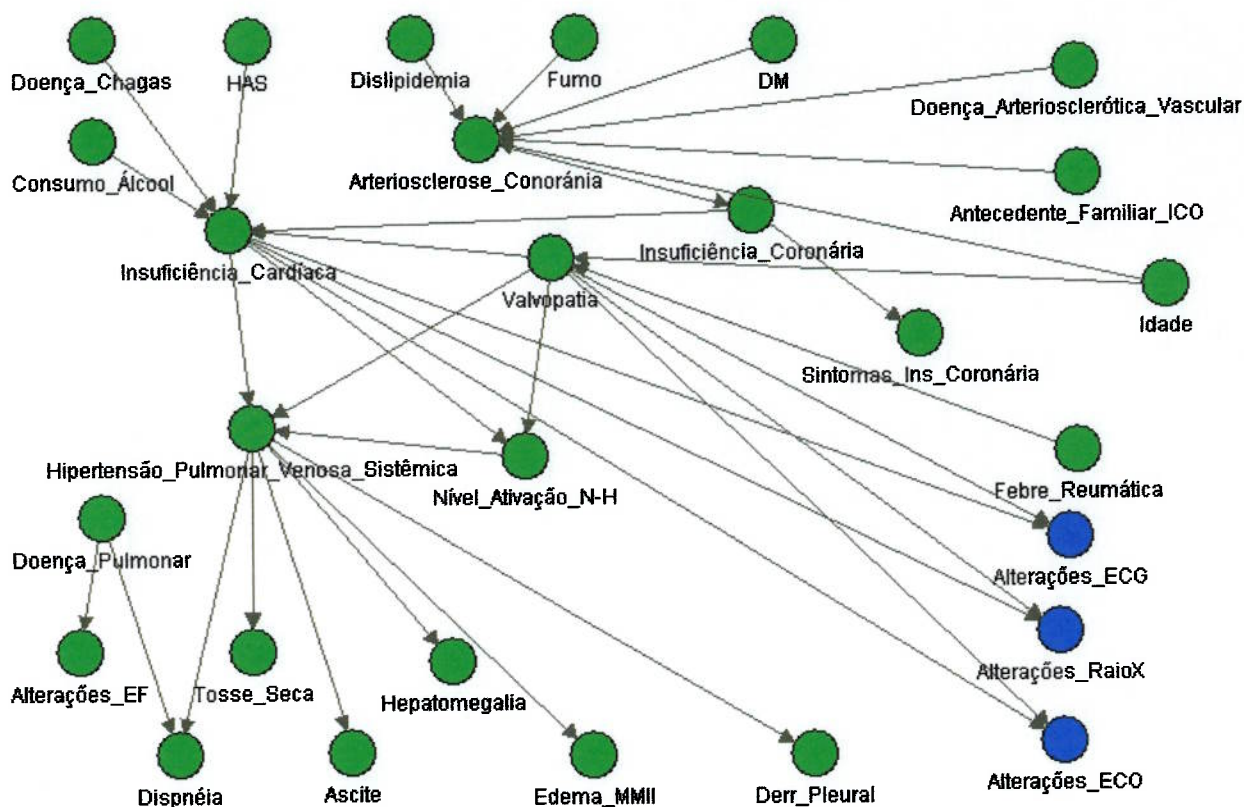


Figura 5.12. Rede Cardiomiopatia.

O caso indicado na Figura 5.12 representa uma situação hipotética em que todos os exames apresentam resultado positivo.

Aplicando os conceitos de d-separação a rede é reduzida a dezessete variáveis. O conjunto de sensibilidade pode ser visto na Figura 5.13.

A árvore de junção é formada por 14 *buckets* e existem várias centenas de probabilidades em estudo. Serão indicados aqui apenas as duas probabilidades em estudo; a primeira influencia a probabilidade de interesse e a segunda quase não possui influência alguma (Figuras 5.14 e 5.15).

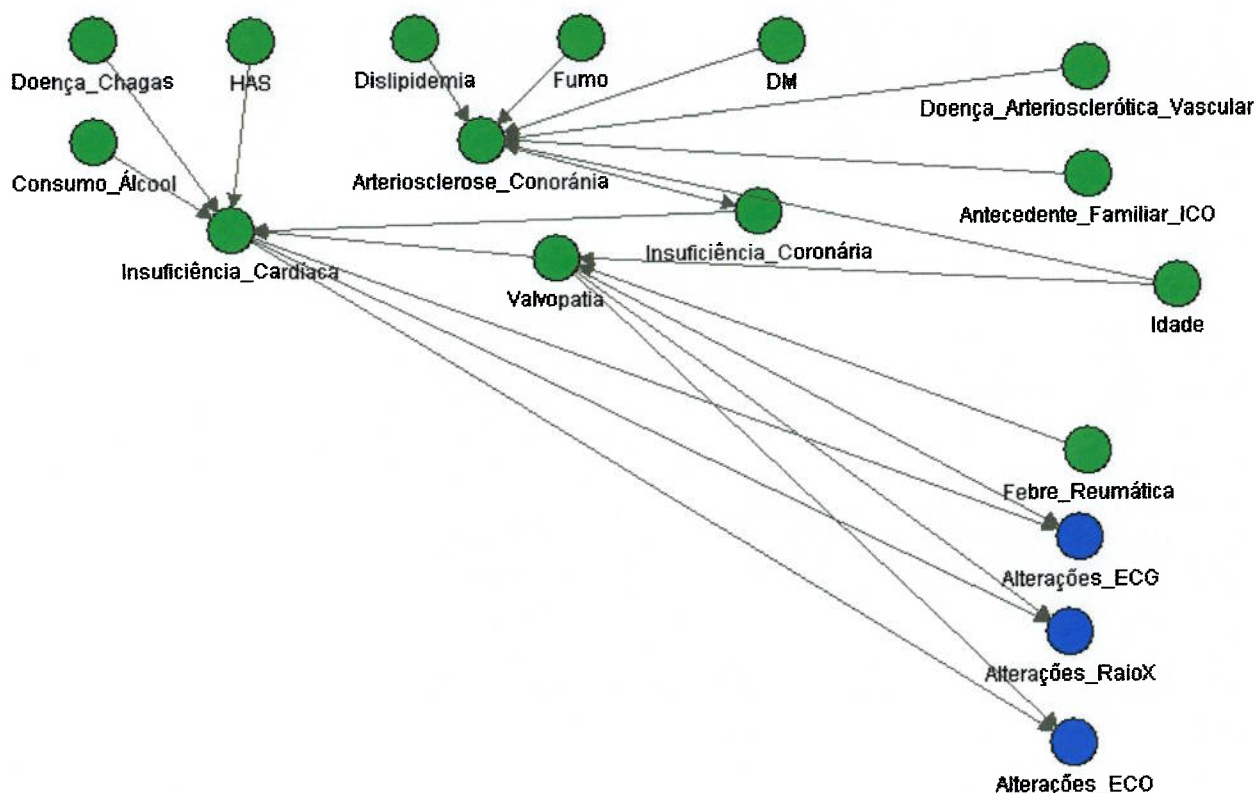


Figura 5.13. Rede Cardiomiopatia - conjunto de sensibilidade.

Na rede Cardiomiopatia acima, com $\{Ins.Card. = Sim\}$ como variável de interesse e $\{Alt.ECG = Sim\}$, $\{Alt.RaioX = Sim\}$ e $\{Alt.ECO = Sim\}$ como evidências e aplicando o conceito de d-separação, temos uma redução de 28 variáveis iniciais para 17 pertencentes ao conjunto de sensibilidade. Essa redução no número de variáveis que devem ser consideradas nos cálculos demonstra os benefícios de se realizar uma análise qualitativa na rede antes de se iniciar a análise quantitativa.

A Tabela 5.3 indica os valores dos coeficientes α , β , γ e δ para as duas variáveis em estudo.

Tabela 5.3. Tabela de coeficientes das duas probabilidades em estudo.

Variável em estudo	α	β	γ	δ
Febre Reumática=Não	-0.35801	1.169611	-0.745257	1.707994
Idade=66-75	0.013422	0.824804	0.005936	0.997922

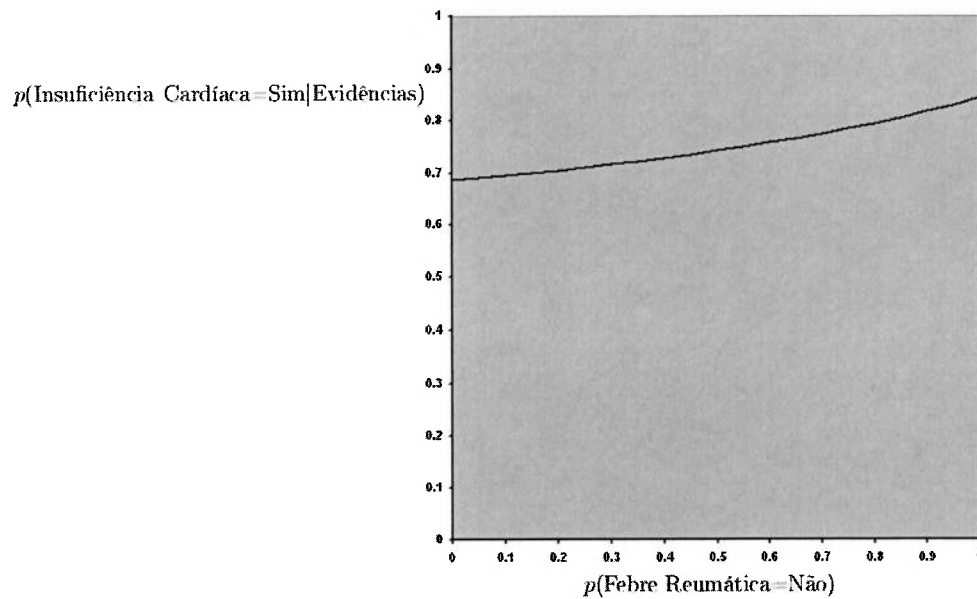


Figura 5.14. Curva de derivada: probabilidade de interesse $p(\text{Ins.Card.}=\text{Sim}|\text{Evidências})$ e probabilidade em estudo $p(\text{Febre Reumática}=\text{Não})$. A curva dessa probabilidade em estudo foi escolhida por estar entre as que mais influenciam o valor da probabilidade de interesse. Mesmo assim a variação na probabilidade de interesse não chega a 20%.

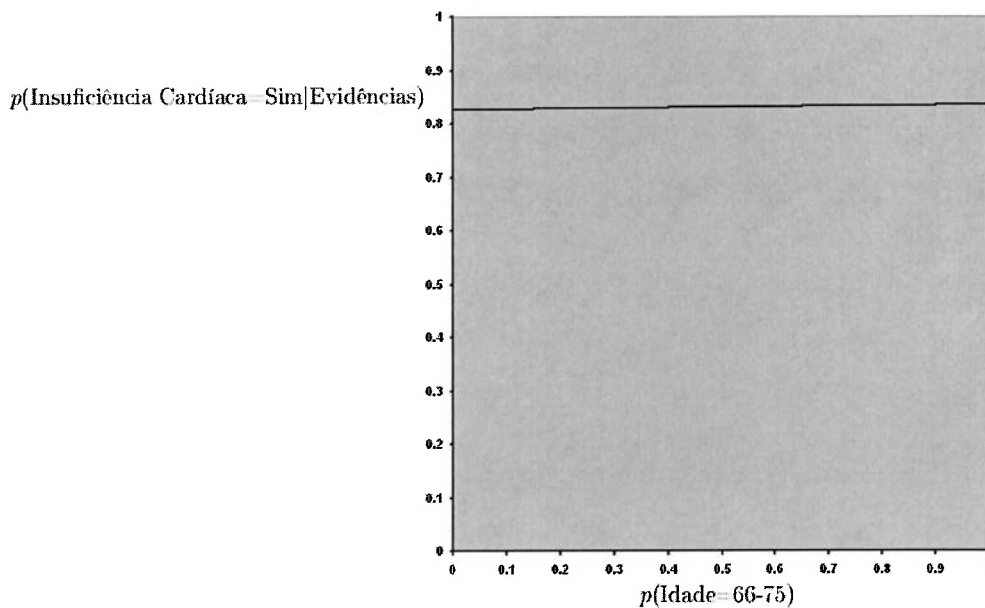


Figura 5.15. Curva de derivada: probabilidade de interesse $p(\text{Ins.Card.}=\text{Sim}|\text{Evidências})$ e probabilidade em estudo $p(\text{Idade}=66-75)$.

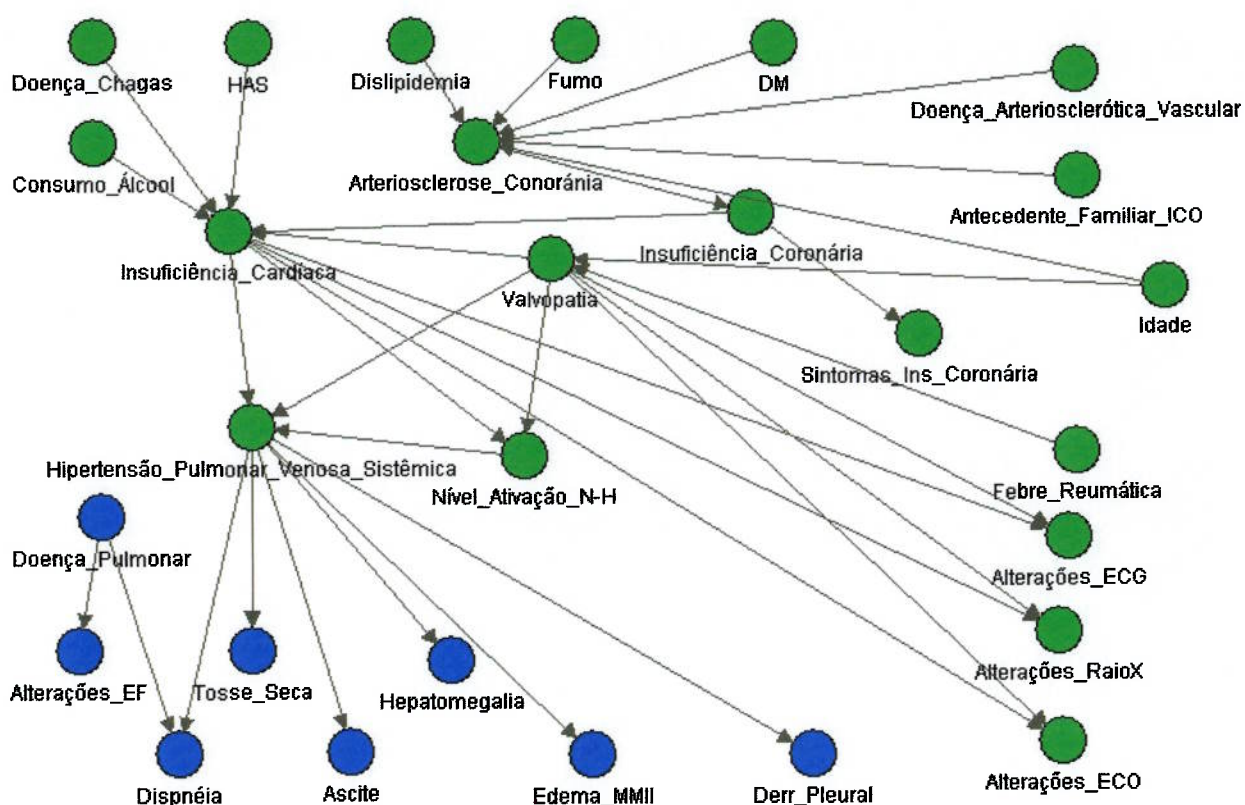


Figura 5.16. Rede Cardiomiopatia 2.

Escolhendo um segundo conjunto de evidências mais resultados foram obtidos para a rede Cardiomiopatia.

A Figura 5.16 mostra a influência dos nós representando a ausência de sintomas do paciente no diagnóstico de insuficiência cardíaca. Apesar da indicação de ausência de sintomas diminua a probabilidade do nó Insuficiência Cardíaca, não se tem uma probabilidade próxima a 0 devido à falta de outras informações, como dados do paciente e resultados de exames diagnósticos [27].

Novamente aplicando os conceitos de d-separação a rede é reduzida a vinte e duas variáveis. O conjunto de sensibilidade pode ser visto na Figura 5.17.

A árvore de junção é formada por 16 *buckets* e existem várias centenas de probabilidades em estudo. Serão indicados aqui apenas as duas probabilidades em estudo cuja variação mais influenciam a probabilidade de interesse (Figuras 5.18 e 5.19).

A Tabela 5.4 indica os valores dos coeficientes α , β , γ e δ para as duas variáveis em estudo neste segundo conjunto de evidências com a rede cardiomiopatia.

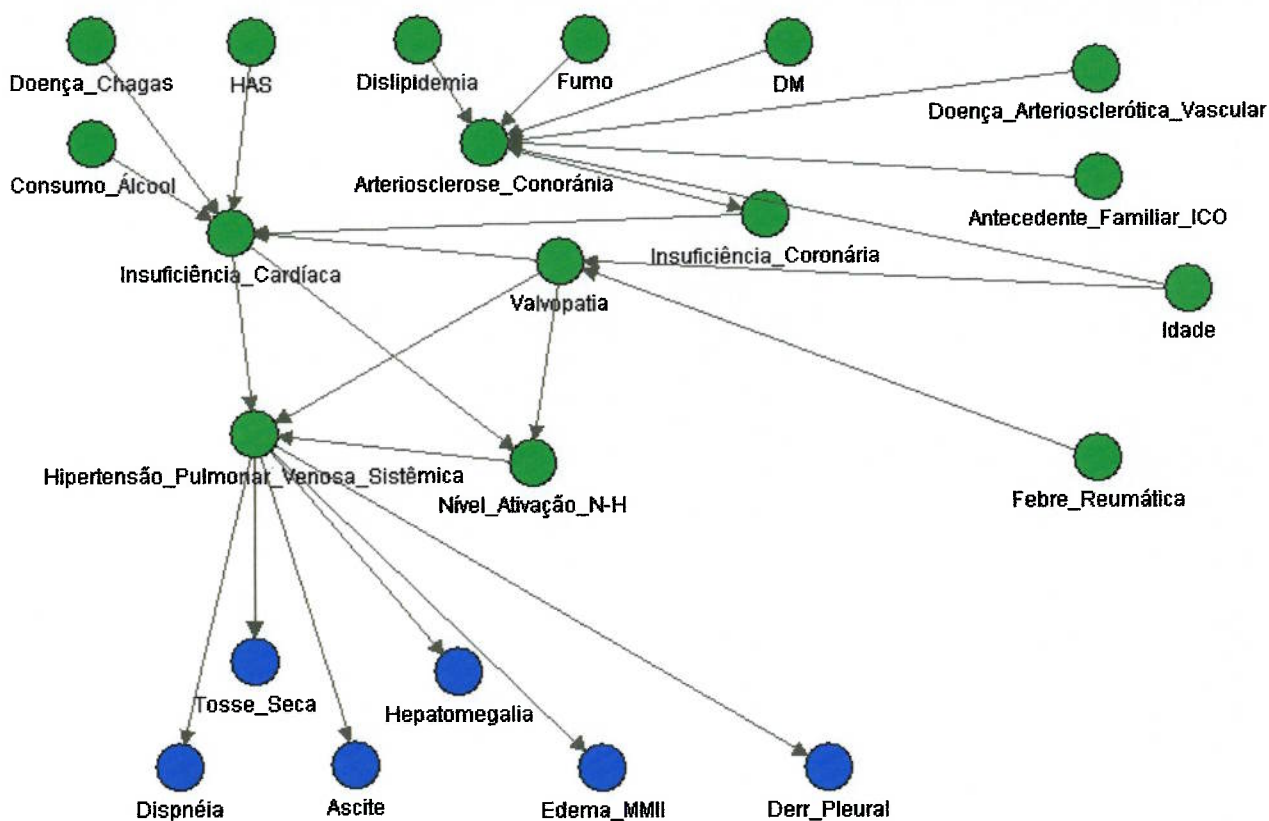


Figura 5.17. Rede Cardiomiopatia - conjunto de sensibilidade 2.

Neste segundo exemplo com a rede Cardiomiopatia como um conjunto diferente de evidências também se reduz o número de variáveis ao se aplicar o conceito de d-separação.

Tabela 5.4. Tabela de coeficientes das duas probabilidades em estudo 2.

Variável em estudo	α	β	γ	δ
Hipertensão Pulmonar Venosa Sistêmica=Não / Valvopatia=Não, Insuficiência Cardíaca=Não, Nível Ativação N-H=Não	0	0.18843	0.606902	0.405236
Doença Chagas=Não	-0.306643	0.47974	0.236102	0.775703

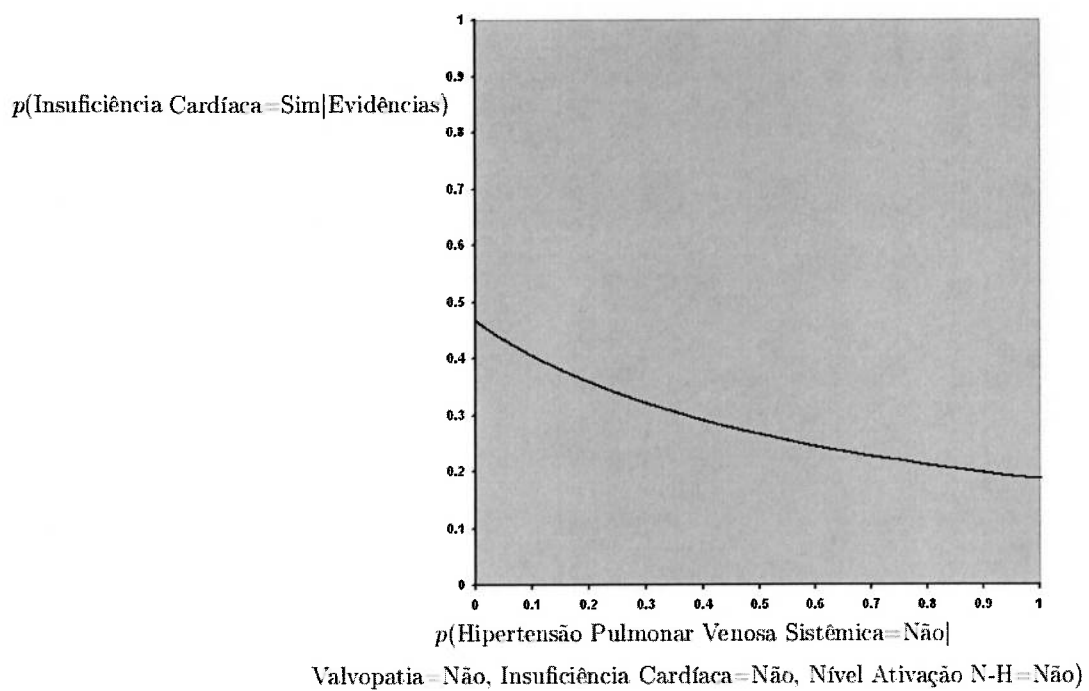


Figura 5.18. Curva de derivada: probabilidade de interesse $p(\text{Ins.Card.}=\text{Sim} | \text{Evidências})$ e probabilidade em estudo $p(\text{Hipertensão Pulmonar Venosa Sistêmica=Não} | \text{Valvopatia=Não, Insuficiência Cardíaca=Não, Nível Ativação N-H=Não})$.

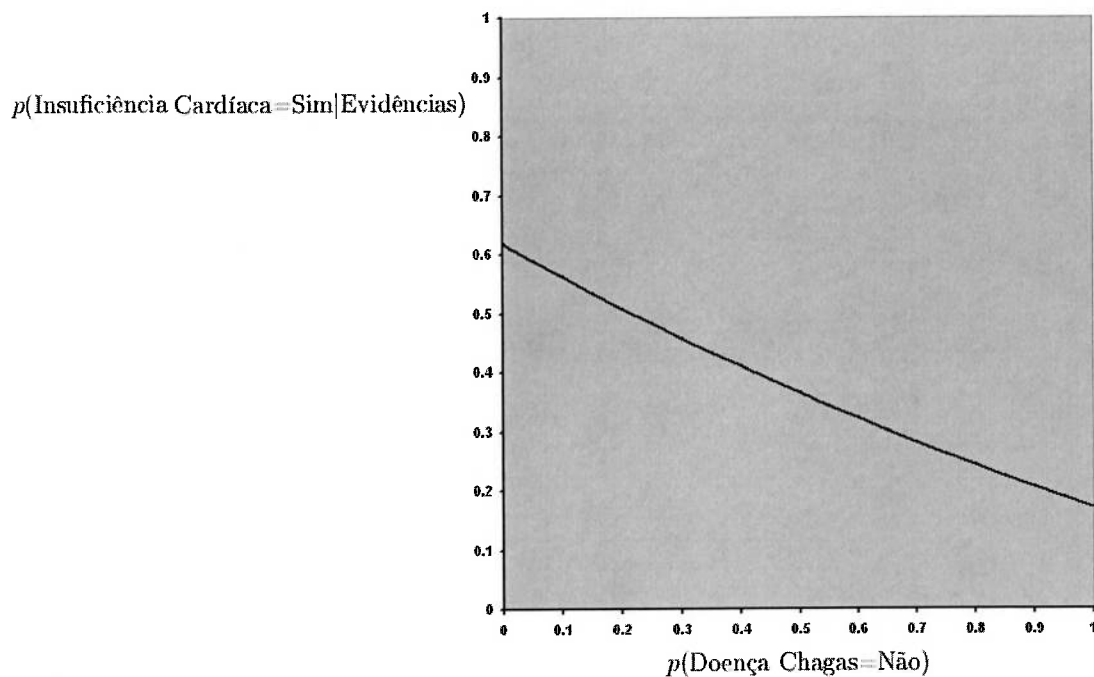


Figura 5.19. Curva de derivada: probabilidade de interesse $p(\text{Ins.Card.}=\text{Sim}|\text{Evidências})$ e probabilidade em estudo $p(\text{Doença Chagas}=\text{Não})$. Nesta segunda configuração de evidências com a rede Cardiomiopatia tem-se probabilidades em estudo com influências mais destacadas do que na configuração anterior. As probabilidades em estudo da figura acima e da Figura 5.18 seriam apontados numa análise de sensibilidade como possíveis candidatos a um refinamento. Porém, antes de se decidir pelo refinamento dessas probabilidades deve-se levar em consideração se o custo dos refinamentos supera os benefícios de uma maior precisão, impossibilidade de se obter um maior refinamento devido a falta de conhecimento.

Capítulo 6

CONCLUSÕES

O algoritmo apresentado neste trabalho para se obter derivadas em redes Bayesianas é de fácil entendimento e implementação. Diferente de métodos atuais, como o apresentado por Kjaerulff e van der Gaag [21], este algoritmo não utiliza conceitos de grafos, tais como triangulação e cliques, evitando dessa forma a utilização de complexos métodos da teoria de grafos. Ao invés disso, este algoritmo se foca apenas em densidades de probabilidades. O Capítulo 4, além de propor o novo método, demonstra seu funcionamento com uma rede Bayesiana de exemplo.

Aplicações deste algoritmo são de grande utilidade. Métodos eficientes para análise de sensibilidade tem uma função importante tanto nas fases de aquisição de conhecimento como na fase de validação na construção manual de modelos de redes Bayesianas. A otimização local de valores de probabilidade tem aplicações quando parâmetros não são conhecidos com precisão absoluta, por serem baseados em conhecimento limitados ou em dados experimentais finitos.

O novo método foi implementado na linguagem de programação Java, como um acréscimo a biblioteca Embayes, adicionando a essa biblioteca a capacidade de derivar redes Bayesianas.

Os resultados apresentados no Capítulo 5 desta dissertação não apenas demonstraram o funcionamento do programa em Java desenvolvido, mas serviram também para mais uma vez validar o novo método proposto para se obter derivadas em redes Bayesianas.

Todos os resultados experimentais obtidos foram validados, seja através de cálculos manuais e variação sistemática para a primeira rede de exemplo de pequeno

tamanho ou através de três inferências na rede (com três valores distintos da respectiva probabilidade em estudo) com base no segundo princípio do método proposto por Coupé *et al* [7].

Espera-se ter apresentado de maneira clara e objetiva o novo método para o cálculo de derivadas de parâmetros de rede Bayesianas assim como os resultados obtidos com a aplicação deste novo método.

Os resultados apresentados nessa dissertação podem ser estendidos em várias direções significativas. Uma direção inicial seria o cálculo da variação de uma probabilidade em relação a n parâmetros simultaneamente, por exemplo para estudar sensibilidade em relação a duas ou mais valores de probabilidade. Outro passo futuro seria a integração dos algoritmos apresentados nesse trabalho com programas de otimização local por método do gradiente, permitindo assim que valores máximos e mínimos de probabilidades sejam calculados de forma eficiente.

Referências Bibliográficas

- [1] S. Andreassen. A model-based approach to insulin adjustment. *Proceedings of the Third Conference on Artificial Intelligence in Medicine, Lecture Notes in Medical Informatics 44*, 239-248, 1991.
- [2] I. Beinlich, H. J. Suermondt, R. M. Chavez, e G. F. Cooper. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. *Second European Conference on Artificial Intelligence in Medicine*, 247-256, 1989.
- [3] 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.
- [4] E. Castillo, J. M. Gutierrez, e A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, New York, 1997.
- [5] H. Chan e A. Darwiche. When do numbers really matter? *Conference on Uncertainty in Artificial Intelligence*, 65-74, 2001.
- [6] G. Chartrand e O. R. Oellermann. *Applied and Algorithmic Graph Theory*. McGraw-Hill, New York, 1993.
- [7] V. M. H. Coupé, L. C. van der Gaag, e J. D. F. Habbema. *Sensitivity Analysis of Decision-Theoretic Networks*. Holanda, 1999.
- [8] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, e D. J. Spiegelhalter. *Probabilistic Networks e Expert Systems*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 1999.
- [9] F. G. Cozman. Credal networks. *Artificial Intelligence*, 199-233, 2000.

- [10] F. G. Cozman. Generalizing variable elimination in Bayesian networks. *Workshop on Probabilistic Reasoning in Artificial Intelligence*, Novembro 2000.
- [11] A. Darwiche e G. Provan. Query DAGs: A practical paradigm for implementing belief-network inference. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI-1996)*, 203-210, San Francisco, Estados Unidos, 1996. Morgan Kaufmann.
- [12] R. Dechter. Bucket elimination: An unifying framework for probabilistic inference. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI-96)*, 211-219, San Francisco, agosto 1996. Morgan Kaufmann.
- [13] S. Dwarkadas. Parallelization of general linkage analysis problems. *Human Heredity*, v44, 127-141, 1994.
- [14] E. Fagiouli e M. Zaffalon. *2U: An exact interval propagation algorithm for polytrees with binary variables*. *Artificial Intelligence* 106(1)77:107. 1998.
- [15] E. Horvitz. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 256-265, 1998.
- [16] E. Horvitz e M. Barry. Display of information for time-critical decision making. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 286-305, 1995.
- [17] J. S. Ide. *Geração de Redes Bayesianas Uniformemente Distribuídas*. Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para a obtenção do Título de Mestre em Engenharia. São Paulo, Brasil, 2002.
- [18] F. V. Jensen. *An Introduction to Bayesian Networks*. Springer Verlag, 1996.
- [19] F. V. Jensen e F. Jensen. Optimal junction trees. *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, 360-366, 1994.
- [20] U. Kjaerulff. *Triangulation of Graphs - Algorithms Giving Small Total State Space*. Departamento de matemática e Ciência da Computação. Universidade de Aalborg, Dinamarca, 1990.

- [21] U. Kjærulff e L. C. van der Gaag. Making sensitivity analysis computationally efficient. *Conference on Uncertainty in Artificial Intelligence*, 2000.
- [22] M. Korver e P. J. F. Lucas. Converting a rulebased expert system into a belief network. *Medical Informatics*, 18:219–241, 1993.
- [23] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1998.
- [24] N. B. Peek e J. Ottenkamp. Developing a decision-theoretic network for a congenital heart disease. *Proceedings of the Sixth Conference on Artificial Intelligence in Medicine Europe*, 157-168, 1997.
- [25] F. T. Ramos. *Inferência Probabilística em Sistemas com Restrições de Tempo e Memória*. Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para a obtenção do Título de Mestre em Engenharia. São Paulo, Brasil, 2003.
- [26] S. J. Russell e P. Norvig. *Artificial Intelligence. A Modern Approach*. Prentice-Hall, Englewood Cliffs, 1995.
- [27] A. H. Saheki. *Construção de uma rede Bayesiana aplicada ao diagnóstico de doenças cardíacas*. Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para a obtenção do Título de Mestre em Engenharia. São Paulo, Brasil, 2005.
- [28] J. Stutz, W. Taylor, e P. Cheeseman. *AutoClass C - General Information*. NASA, Ames Research Center. 1998.
- [29] N. L. Zhang e D. Poole. *Exploiting causal independence in Bayesian network inference*. *Journal of Artificial Intelligence Research*, 301-328. 1996.
- [30] G. Zweig e S. J. Russel. Speech recognition with dynamic Bayesian networks. *Proceedings of the Fifteenth National Conference on Artificial Intelligence AAAI/IAAI*, 173-180, 1998.

APÊNDICES

Apêndice A

ARQUIVOS DE ENTRADA

A.1 Rede de exemplo (5 nós)

```
// Bayesian network in BIF format
```

```
network "cba2004" { //5 nodes }
```

```
variable "C" { //2 categories
```

```
    type discrete[2] { "true" "false" };
```

```
    property "position = (150, 250)";
```

```
} variable "B" { //2 categories
```

```
    type discrete[2] { "true" "false" };
```

```
    property "position = (50, 250)";
```

```
} variable "A" { //2 categories
```

```
    type discrete[2] { "true" "false" };
```

```
    property "position = (100, 200)";
```

```
} variable "E" { //2 categories
```

```
    type discrete[2] { "true" "false" };
```

```
    property "position = (200, 300)";
```

```
} variable "D" { //2 categories
```

```
    type discrete[2] { "true" "false" };
```

```
    property "position = (100, 300)";
```

```
} probability ( "C" "A" ) { //2 variable(s) and 4 values
```

```

    table 0.2 0.05 0.8 0.95;
} probability ( "B" "A" ) { //2 variable(s) and 4 values
    table 0.8 0.2 0.2 0.8;
}
probability ( "A" ) { //1 variable(s) and 2 values
    table 0.2 0.8;
}
probability ( "E" "C" ) { //2 variable(s) and 4 values
    table 0.8 0.6 0.2 0.4;
}
probability ( "D" "B" "C" ) { //3 variable(s) and 8 values
    table 0.8 0.8 0.8 0.05 0.2 0.2 0.2 0.95;
}

```

A.2 Rede Alarm (37 nós)

```

// Bayesian Network in the Interchange Format
// Produced by BayesianNetworks package in JavaBayes
// Output created Sun Nov 02 17:58:15 GMT+00:00 1997
// Bayesian network
network "Alarm" { //37 variables and 37 probability distributions
}
variable "Hypovolemia" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (54, 35)" ;
}
variable "StrokeVolume" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (184, 113)" ;
}
variable "LVFailure" { //2 values

```



```
    type discrete[2] { "True" "False" };
    property "position = (145, 36)" ;
}
variable "LVEDVolume" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (68, 114)" ;
}
variable "PCWP" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (111, 177)" ;
}
variable "CVP" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (32, 179)" ;
}
variable "History" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (238, 61)" ;
}
variable "MinVolSet" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (564, 38)" ;
}
variable "VentMach" { //4 values
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (640, 86)" ;
}
variable "Disconnect" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (738, 86)" ;
}
variable "VentTube" { //4 values
```

```
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (682, 168)" ;
}
variable "KinkedTube" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (564, 172)" ;
}
variable "Press" { //4 values
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (722, 253)" ;
}
variable "ErrLowOutput" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (226, 237)" ;
}
variable "HRBP" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (229, 305)" ;
}
variable "ErrCauter" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (366, 278)" ;
}
variable "HREKG" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (289, 305)" ;
}
variable "HRSat" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (220, 396)" ;
}
variable "BP" { //3 values
```

```
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (154, 396)" ;
}
variable "CO" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (195, 176)" ;
}
variable "HR" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (308, 171)" ;
}
variable "TPR" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (120, 301)" ;
}
variable "Anaphylaxis" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (31, 239)" ;
}
variable "InsuffAnesth" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (329, 37)" ;
}
variable "PAP" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (1045, 292)" ;
}
variable "PulmEmbolus" { //2 values
    type discrete[2] { "True" "False" };
    property "position = (969, 258)" ;
}
variable "FiO2" { //2 values
```

```
    type discrete[2] { "Low" "Normal" };
    property "position = (1014, 162)" ;
}
variable "Catechol" { //2 values
    type discrete[2] { "Normal" "High" };
    property "position = (329, 107)" ;
}
variable "SaO2" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (926, 387)" ;
}
variable "Shunt" { //2 values
    type discrete[2] { "Normal" "High" };
    property "position = (894, 293)" ;
}
variable "PVSat" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (949, 197)" ;
}
variable "MinVol" { //4 values
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (754, 387)" ;
}
variable "ExpCO2" { //4 values
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (530, 393)" ;
}
variable "ArtCO2" { //3 values
    type discrete[3] { "Low" "Normal" "High" };
    property "position = (474, 277)" ;
}
variable "VentAlv" { //4 values
```

```
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (881, 165)" ;
}
variable "VentLung" { //4 values
    type discrete[4] { "Zero" "Low" "Normal" "High" };
    property "position = (706, 344)" ;
}
variable "Intubation" { //3 values
    type discrete[3] { "Normal" "Esophageal" "OneSided" };
    property "position = (843, 86)" ;
}
probability ( "Hypovolemia" ) { //1 variable(s) and 2 values
    table 0.2 0.8 ;
}
probability ( "StrokeVolume" "LVFailure" "Hypovolemia" ) {
//3 variable(s) and 12 values
    table 0.98 0.5 0.95 0.05 0.01 0.49 0.04 0.9 0.01 0.01 0.01 0.05 ;
}
probability ( "LVFailure" ) { //1 variable(s) and 2 values
    table 0.05 0.95 ;
}
probability ( "LVEDVolume" "Hypovolemia" "LVFailure" ) { //3
variable(s) and 12 values
    table 0.95 0.98 0.01 0.05 0.04 0.01 0.09 0.9 0.01 0.01 0.9 0.05 ;
}
probability ( "PCWP" "LVEDVolume" ) { //2 variable(s) and 9
values
    table 0.95 0.04 0.01 0.04 0.95 0.04 0.01 0.01 0.95 ;
}
probability ( "CVP" "LVEDVolume" ) { //2 variable(s) and 9
values
    table 0.95 0.04 0.01 0.04 0.95 0.29 0.01 0.01 0.7 ;
```

```
}
probability ( "History" "LVFailure" ) { //2 variable(s) and 4
values
    table 0.9 0.01 0.1 0.99 ;
}
probability ( "MinVolSet" ) { //1 variable(s) and 3 values
    table 0.01 0.98 0.01 ;
}
probability ( "VentMach" "MinVolSet" ) { //2 variable(s) and
12 values
    table 0.01 0.01 0.01 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01
0.97 ;
}
probability ( "Disconnect" ) { //1 variable(s) and 2 values
    table 0.05 0.95 ;
}
probability ( "VentTube" "VentMach" "Disconnect" ) { //3
variable(s) and 32 values
    table 0.97 0.97 0.97 0.01 0.97 0.01 0.97 0.01 0.01 0.01 0.01
0.97 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.97 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.97 ;
}
probability ( "KinkedTube" ) { //1 variable(s) and 2 values
    table 0.04 0.96 ;
}
probability ( "Press" "KinkedTube" "Intubation" "VentTube" )
{ //4 variable(s) and 96 values
    table 0.97 0.01 0.01 0.01 0.97 0.1 0.05 0.01 0.97 0.01 0.01
0.01 0.97 0.01 0.01 0.01 0.97 0.4 0.2 0.2 0.97 0.01 0.01 0.01
0.01 0.49 0.01 0.01 0.01 0.84 0.25 0.15 0.01 0.29 0.01 0.01
0.01 0.97 0.01 0.01 0.01 0.58 0.75 0.7 0.01 0.9 0.01 0.01 0.01
0.3 0.08 0.01 0.01 0.05 0.25 0.25 0.01 0.3 0.08 0.01 0.01 0.01
```

```
0.97 0.01 0.01 0.01 0.04 0.09 0.01 0.08 0.38 0.01 0.01 0.2 0.9
0.97 0.01 0.01 0.45 0.59 0.01 0.4 0.9 0.97 0.01 0.01 0.01 0.97
0.01 0.01 0.01 0.01 0.01 0.01 0.6 0.97 ;
}
probability ( "ErrLowOutput" ) { //1 variable(s) and 2 values
    table 0.05 0.95 ;
}
probability ( "HRBP" "ErrLowOutput" "HR" ) { //3 variable(s)
and 18 values
    table 0.98 0.4 0.3 0.98 0.01 0.01 0.01 0.59 0.4 0.01 0.98 0.01
    0.01 0.01 0.3 0.01 0.01 0.98 ;
}
probability ( "ErrCauter" ) { //1 variable(s) and 2 values
    table 0.1 0.9 ;
}
probability ( "HREKG" "HR" "ErrCauter" ) { //3 variable(s)
and 18 values
    table 0.333333 0.98 0.333333 0.01 0.333333 0.01 0.333333 0.01
    0.333333 0.98 0.333333 0.01 0.333333 0.01 0.333333 0.01
    0.333333 0.98 ;
}
probability ( "HRSat" "HR" "ErrCauter" ) { //3 variable(s)
and 18 values
    table 0.333333 0.98 0.333333 0.01 0.333333 0.01 0.333333 0.01
    0.333333 0.98 0.333333 0.01 0.333333 0.01 0.333333 0.01
    0.333333 0.98 ;
}
probability ( "BP" "CO" "TPR" ) { //3 variable(s) and 27
values
    table 0.98 0.98 0.3 0.98 0.1 0.05 0.9 0.05 0.01 0.01 0.01 0.6
    0.01 0.85 0.4 0.09 0.2 0.09 0.01 0.01 0.1 0.01 0.05 0.55 0.01
    0.75 0.9 ;
```

```
}
probability ( "CO" "HR" "StrokeVolume" ) { //3 variable(s)
and 27 values
    table 0.98 0.95 0.3 0.95 0.04 0.01 0.8 0.01 0.01 0.01 0.04 0.69
    0.04 0.95 0.3 0.19 0.04 0.01 0.01 0.01 0.01 0.01 0.01 0.69 0.01
    0.95 0.98 ;
}
probability ( "HR" "Catechol" ) { //2 variable(s) and 6 values
    table 0.1 0.01 0.89 0.09 0.01 0.9 ;
}
probability ( "TPR" "Anaphylaxis" ) { //2 variable(s) and 6
values
    table 0.98 0.3 0.01 0.4 0.01 0.3 ;
}
probability ( "Anaphylaxis" ) { //1 variable(s) and 2 values
    table 0.01 0.99 ;
}
probability ( "InsuffAnesth" ) { //1 variable(s) and 2 values
    table 0.2 0.8 ;
}
probability ( "PAP" "PulmEmbolus" ) { //2 variable(s) and 6
values
    table 0.01 0.05 0.19 0.9 0.8 0.05 ;
}
probability ( "PulmEmbolus" ) { //1 variable(s) and 2 values
    table 0.01 0.99 ;
}
probability ( "FiO2" ) { //1 variable(s) and 2 values
    table 0.01 0.99 ;
}
probability ( "Catechol" "InsuffAnesth" "SaO2" "TPR"
"ArtCO2" ) { //5 variable(s) and 108 values
```



```

table 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.05 0.05 0.01 0.01 0.01 0.01 0.05 0.05 0.01 0.05
0.05 0.01 0.05 0.05 0.01 0.05 0.05 0.01 0.05 0.05 0.01 0.1 0.1
0.1 0.95 0.95 0.3 0.95 0.95 0.3 0.95 0.95 0.3 0.99 0.99 0.99 0.95
0.99 0.3 0.99 0.99 0.99 0.99 0.99 0.99 0.99 0.99 0.99 0.99 0.99
0.99 0.99 0.99 0.99 0.95 0.95 0.99 0.99 0.99 0.99 0.95 0.95 0.99
0.95 0.95 0.99 0.95 0.95 0.99 0.95 0.95 0.99 0.95 0.95 0.99 0.9
0.9 0.9 0.05 0.05 0.7 0.05 0.05 0.7 0.05 0.05 0.7 0.01 0.01 0.01
0.05 0.01 0.7 ;
}
probability ( "SaO2" "Shunt" "PVSat" ) { //3 variable(s) and
18 values
table 0.98 0.01 0.01 0.98 0.98 0.69 0.01 0.98 0.01 0.01 0.01 0.3
0.01 0.01 0.98 0.01 0.01 0.01 ;
}
probability ( "Shunt" "PulmEmbolus" "Intubation" ) { //3
variable(s) and 12 values
table 0.1 0.1 0.01 0.95 0.95 0.05 0.9 0.9 0.99 0.05 0.05 0.95 ;
}
probability ( "PVSat" "VentAlv" "FiO2" ) { //3 variable(s)
and 24 values
table 0.98 0.98 0.98 0.98 0.95 0.01 0.95 0.01 0.01 0.01 0.01 0.01
0.04 0.95 0.04 0.01 0.01 0.01 0.01 0.01 0.01 0.04 0.01 0.98 ;
}
probability ( "MinVol" "VentLung" "Intubation" ) { //3
variable(s) and 48 values
table 0.97 0.97 0.97 0.01 0.6 0.01 0.01 0.5 0.01 0.01 0.5 0.01
0.01 0.01 0.01 0.97 0.38 0.97 0.01 0.48 0.01 0.01 0.48 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.97 0.01 0.97 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.97 0.01 0.97 ;
}
probability ( "ExpCO2" "ArtCO2" "VentLung" ) { //3

```

variable(s) and 48 values

```
table 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01
0.01 0.97 0.97 0.97 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.97 0.97 0.97 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.97 0.97 0.97 ;
```

}

probability ("ArtCO2" "VentAlv") { //2 variable(s) and 12

values

```
table 0.01 0.01 0.04 0.9 0.01 0.01 0.92 0.09 0.98 0.98 0.04 0.01 ;
```

}

probability ("VentAlv" "Intubation" "VentLung") { //3

variable(s) and 48 values

```
table 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01 0.97 0.03 0.01 0.01
0.01 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01 0.95 0.94 0.88 0.01
0.01 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01 0.04 0.1 0.01 0.01
0.01 0.97 0.01 0.01 0.01 0.97 0.01 0.01 0.01 0.01 ;
```

}

probability ("VentLung" "KinkedTube" "VentTube"

"Intubation") { //4 variable(s) and 96 values

```
table 0.97 0.97 0.97 0.95 0.97 0.95 0.4 0.97 0.5 0.3 0.97 0.3
0.97 0.97 0.97 0.01 0.97 0.01 0.01 0.97 0.01 0.01 0.97 0.01 0.01
0.01 0.01 0.03 0.01 0.03 0.58 0.01 0.48 0.68 0.01 0.68 0.01 0.01
0.01 0.97 0.01 0.97 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.97 0.01 0.97 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.97 0.01 0.97 ;
```

}

probability ("Intubation") { //1 variable(s) and 3 values

```
table 0.92 0.03 0.05 ;
```

}

A.3 Rede Cardiomiopatia (28 nós)

```
// Bayesian network in BIF format

network "BayesianNetwork" { //28 nodes } variable "Edema_MMII" {
//2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (342, 435)";
} variable "Alterações_EF" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (43, 374)";
} variable "Valvopatia" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (319, 146)";
} variable "Doença_Chagas" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (52, 18)";
} variable "HAS" { //5 categories
    type discrete[5] { "Normal" "Normal_Alta" "HT_Baixa"
"HT_Média" "HT_Elevada" };
    property "position = (143, 19)";
} variable "Alterações_ECO" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (617, 428)";
} variable "Dislipidemia" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (238, 18)";
} variable "Alterações_ECG" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (621, 309)";
} variable "Alterações_RaioX" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (615, 364)";
```

```
} variable "Doença_Pulmonar" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (56, 296)";
} variable "Ascite" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (201, 433)";
} variable "Idade" { //5 categories
  type discrete[5] { "<41" "41-55" "56-65" "66-75" ">75" };
  property "position = (678, 163)";
} variable "DM" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (450, 20)";
} variable "Nível_Ativação_N-H" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (303, 261)";
} variable "Insuficiência_Coronária" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (436, 119)";
} variable "Arteriosclerose_Coronária" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (276, 77)";
} variable "Sintomas_Ins_Coronária" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (534, 191)";
} variable "Febre_Reumática" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (626, 259)";
} variable "Tosse_Seca" { //2 categories
  type discrete[2] { "Não" "Sim" };
  property "position = (144, 374)";
} variable "Consumo_Álcool" { //2 categories
  type discrete[2] { "Não" "Sim" };
```

```
    property "position = (52, 77)";
} variable "Insuficiência_Cardíaca" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (131, 129)";
} variable "Hepatomegalia" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (260, 380)";
} variable "Antecedente_Familiar_ICO" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (627, 98)";
} variable "Dispnéia" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (101, 434)";
} variable "Derr_Pleural" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (471, 435)";
} variable "Hipertensão_Pulmonar_Venosa_Sistêmica" { //2
categories
    type discrete[2] { "Não" "Sim" };
    property "position = (144, 244)";
} variable "Doença_Arteriosclerótica_Vascular" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (626, 36)";
} variable "Fumo" { //2 categories
    type discrete[2] { "Não" "Sim" };
    property "position = (334, 18)";
}

probability ( "Edema_MMII"
"Hipertensão_Pulmonar_Venosa_Sistêmica" ) { //2 variable(s) and 4
values
    table 0.75 0.6 0.25 0.4;
```

```
} probability ( "Alterações_EF" "Doença_Pulmonar" ) { //2
variable(s) and 4 values
    table 0.97 0.2 0.03 0.8;
} probability ( "Valvopatia" "Febre_Reumática" "Idade" ) { //3
variable(s) and 20 values
    table 0.98 0.97 0.94 0.85 0.7 0.3 0.32 0.36 0.42 0.6 0.02 0.03
    0.06 0.15 0.3 0.7 0.68 0.64 0.58 0.4;
} probability ( "Doença_Chagas" ) { //1 variable(s) and 2 values
    table 0.95 0.05;
} probability ( "HAS" ) { //1 variable(s) and 5 values
    table 0.4 0.25 0.2 0.1 0.05;
} probability ( "Alterações_ECO" "Valvopatia"
"Insuficiência_Cardíaca" ) { //3 variable(s) and 8 values
    table 0.98 0.02 0.01 0.0010 0.02 0.98 0.99 0.999;
} probability ( "Dislipidemia" ) { //1 variable(s) and 2 values
    table 0.4 0.6;
} probability ( "Alterações_ECG" "Valvopatia"
"Insuficiência_Cardíaca" ) { //3 variable(s) and 8 values
    table 0.7 0.2 0.3 0.1 0.3 0.8 0.7 0.9;
} probability ( "Alterações_RaioX" "Valvopatia"
"Insuficiência_Cardíaca" ) { //3 variable(s) and 8 values
    table 0.9 0.3 0.4 0.2 0.1 0.7 0.6 0.8;
} probability ( "Doença_Pulmonar" ) { //1 variable(s) and 2
values
    table 0.9 0.1;
} probability ( "Ascite" "Hipertensão_Pulmonar_Venosa_Sistêmica"
) { //2 variable(s) and 4 values
    table 0.99 0.9 0.01 0.1;
} probability ( "Idade" ) { //1 variable(s) and 5 values
    table 0.05 0.1 0.2 0.35 0.3;
} probability ( "DM" ) { //1 variable(s) and 2 values
    table 0.7 0.3;
```

```
} probability ( "Nível_Ativação_N-H" "Valvopatia"
"Insuficiência_Cardíaca" ) { //3 variable(s) and 8 values
    table 0.98 0.5 0.5 0.15 0.02 0.5 0.5 0.85;
} probability ( "Insuficiência_Coronária"
"Arteriosclerose_Conorânia" ) { //2 variable(s) and 4 values
    table 0.98 0.0 0.02 1.0;
} probability ( "Arteriosclerose_Conorânia" "Dislipidemia"
"Fumo" "DM" "Antecedente_Familiar_ICO"
"Doença_Arteriosclerótica_Vascular" "Idade" ) { //7 variable(s)
and 320 values
    table 0.95 0.92 0.85 0.65 0.4999 0.497 0.4813 0.4447 0.3401 0.2616
0.8096 0.784 0.7244 0.5539 0.4261 0.4236 0.4102 0.379 0.2898 0.2229
0.6195 0.6 0.5543 0.4239 0.326 0.3241 0.3139 0.29 0.2218 0.1706
0.528 0.5113 0.4724 0.3612 0.2779 0.2762 0.2675 0.2472 0.189 0.1454
0.6861 0.6644 0.6138 0.4694 0.3611 0.359 0.3476 0.3212 0.2456
0.1889 0.5847 0.5662 0.5232 0.4 0.3077 0.3059 0.2963 0.2737 0.2093
0.161 0.4474 0.4333 0.4003 0.3061 0.2355 0.2341 0.2267 0.2094
0.1601 0.1232 0.3813 0.3693 0.3412 0.2609 0.2007 0.1995 0.1932
0.1785 0.1365 0.105 0.7388 0.7155 0.6611 0.5055 0.3888 0.3866
0.3744 0.3459 0.2645 0.2034 0.6297 0.6098 0.5634 0.4308 0.3314
0.3295 0.3191 0.2948 0.2254 0.1734 0.4818 0.4666 0.4311 0.3297
0.2536 0.2521 0.2441 0.2256 0.1725 0.1327 0.4106 0.3977 0.3674
0.281 0.2161 0.2148 0.2081 0.1922 0.147 0.1131 0.5336 0.5167 0.4774
0.3651 0.2808 0.2792 0.2704 0.2498 0.191 0.1469 0.4548 0.4404
0.4069 0.3111 0.2393 0.2379 0.2304 0.2129 0.1628 0.1252 0.348
0.337 0.3113 0.2381 0.1831 0.1821 0.1763 0.1629 0.1245 0.0958
0.2966 0.2872 0.2653 0.2029 0.1561 0.1552 0.1503 0.1388 0.1061
0.0816 0.05 0.08 0.15 0.35 0.5001 0.503 0.5187 0.5553 0.6599
0.7384 0.1904 0.216 0.2756 0.4461 0.5739 0.5764 0.5898 0.621
0.7102 0.7771 0.3805 0.4 0.4457 0.5761 0.674 0.6759 0.6861 0.71
0.7782 0.8294 0.472 0.4887 0.5276 0.6388 0.7221 0.7238 0.7325
0.7528 0.811 0.8546 0.3139 0.3356 0.3862 0.5306 0.6389 0.641
```

```
0.6524 0.6788 0.7544 0.8111 0.4153 0.4338 0.4768 0.6 0.6923 0.6941
0.7037 0.7263 0.7907 0.839 0.5526 0.5667 0.5997 0.6939 0.7645
0.7659 0.7733 0.7906 0.8399 0.8768 0.6187 0.6307 0.6588 0.7391
0.7993 0.8005 0.8068 0.8215 0.8635 0.895 0.2612 0.2845 0.3389
0.4945 0.6112 0.6134 0.6256 0.6541 0.7355 0.7966 0.3703 0.3902
0.4366 0.5692 0.6686 0.6705 0.6809 0.7052 0.7746 0.8266 0.5182
0.5334 0.5689 0.6703 0.7464 0.7479 0.7559 0.7744 0.8275 0.8673
0.5894 0.6023 0.6326 0.719 0.7839 0.7852 0.7919 0.8078 0.853
0.8869 0.4664 0.4833 0.5226 0.6349 0.7192 0.7208 0.7296 0.7502
0.809 0.8531 0.5452 0.5596 0.5931 0.6889 0.7607 0.7621 0.7696
0.7871 0.8372 0.8748 0.652 0.663 0.6887 0.7619 0.8169 0.8179
0.8237 0.8371 0.8755 0.9042 0.7034 0.7128 0.7347 0.7971 0.8439
0.8448 0.8497 0.8612 0.8939 0.9184;
} probability ( "Sintomas_Ins_Coronária"
"Insuficiência_Coronária" ) { //2 variable(s) and 4 values
    table 0.9 0.6 0.1 0.4;
} probability ( "Febre_Reumática" ) { //1 variable(s) and 2
values
    table 0.95 0.05;
} probability ( "Tosse_Seca"
"Hipertensão_Pulmonar_Venosa_Sistêmica" ) { //2 variable(s) and 4
values
    table 0.85 0.5 0.15 0.5;
} probability ( "Insuficiência_Cardíaca" "HAS" "Doença_Chagas"
"Consumo_Álcool" "Insuficiência_Coronária" "Valvopatia" ) { //6
variable(s) and 160 values
    table 0.95 0.6705 0.6861 0.4843 0.8109 0.5724 0.5857 0.4134 0.3352
0.2366 0.2421 0.1709 0.2862 0.202 0.2067 0.1459 0.9394 0.6631
0.6784 0.4789 0.8019 0.566 0.5791 0.4088 0.3315 0.234 0.2394 0.169
0.283 0.1997 0.2044 0.1442 0.8444 0.596 0.6098 0.4305 0.7208
0.5088 0.5206 0.3675 0.298 0.2103 0.2152 0.1519 0.2544 0.1795
0.1837 0.1297 0.7916 0.5588 0.5717 0.4035 0.6758 0.477 0.488
```



```
0.3445 0.2794 0.1972 0.2017 0.1424 0.2385 0.1683 0.1722 0.1215
0.7388 0.5215 0.5336 0.3766 0.6307 0.4452 0.4555 0.3215 0.2607
0.184 0.1883 0.1329 0.2226 0.1571 0.1607 0.1134 0.05 0.3295 0.3139
0.5157 0.1891 0.4276 0.4143 0.5866 0.6648 0.7634 0.7579 0.8291
0.7138 0.798 0.7933 0.8541 0.0606 0.3369 0.3216 0.5211 0.1981
0.434 0.4209 0.5912 0.6685 0.766 0.7606 0.831 0.717 0.8003 0.7956
0.8558 0.1556 0.404 0.3902 0.5695 0.2792 0.4912 0.4794 0.6325
0.702 0.7897 0.7848 0.8481 0.7456 0.8205 0.8163 0.8703 0.2084
0.4412 0.4283 0.5965 0.3242 0.523 0.512 0.6555 0.7206 0.8028
0.7983 0.8576 0.7615 0.8317 0.8278 0.8785 0.2612 0.4785 0.4664
0.6234 0.3693 0.5548 0.5445 0.6785 0.7393 0.816 0.8117 0.8671
0.7774 0.8429 0.8393 0.8866;
} probability ( "Consumo_Álcool" ) { //1 variable(s) and 2 values
    table 0.9 0.1;
} probability ( "Hepatomegalia"
"Hipertensão_Pulmonar_Venosa_Sistêmica" ) { //2 variable(s) and 4
values
    table 0.99 0.85 0.01 0.15;
} probability ( "Antecedente_Familiar_ICO" ) { //1 variable(s)
and 2 values
    table 0.7 0.3;
} probability ( "Dispneia"
"Hipertensão_Pulmonar_Venosa_Sistêmica" "Doença_Pulmonar" ) { //3
variable(s) and 8 values
    table 0.7 0.5 0.4 0.1 0.3 0.5 0.6 0.9;
} probability ( "Derr_Pleural"
"Hipertensão_Pulmonar_Venosa_Sistêmica" ) { //2 variable(s) and 4
values
    table 0.99 0.9 0.01 0.1;
} probability ( "Hipertensão_Pulmonar_Venosa_Sistêmica"
"Valvopatia" "Insuficiência_Cardíaca" "Nível_Ativação_N-H" ) {
//4 variable(s) and 16 values
```

```
    table 0.98 0.2 0.85 0.05 0.9 0.1 0.8 0.01 0.02 0.8 0.15 0.95 0.1
    0.9 0.2 0.99;
} probability ( "Doença_Arteriosclerótica_Vascular" ) { //1
variable(s) and 2 values
    table 0.95 0.05;
} probability ( "Fumo" ) { //1 variable(s) and 2 values
    table 0.8 0.2;
}
```