Incerteza

Capítulo 13 – Russell & Norvig

Seções 13.1 a 13.4

Incerteza

- Seja a ação A_t = sair para o aeroporto t minutos antes do vôo.
- A_t me levará ao aeroporto a tempo?
- Dificuldades de saber o resultado da ação:
  - Estados parcialmente observáveis
    - Estados das estradas, trânsito, etc.
  - Sensores ruidosos
    - Relatórios de trânsito
  - Incerteza quanto ao efeito das ações
    - Acidentes, pneu furado, etc.
  - Grande complexidade em prever e modelar o trânsito

Incerteza

- Um procedimento puramente lógico não é muito útil nesse caso, porque:

1. Arriscaria deduzir algo potencialmente falso
   - "A_{45} me levará a tempo ao aeroporto"
2. Levaria a conclusões fracas para tomada de decisões
   - "A_{45} me levará a tempo ao aeroporto, se nenhum acidente ocorrer na ponte, se não chover, se nenhum pneu furar, etc."
3. Levaria a conclusões que não práticas
   1. "A_{1440} me levará a tempo ao aeroporto"

Lidando com a incerteza

- Probabilidade
  - Modela o **grau de crença** de um agente dadas as evidências disponíveis
    - "A_{25} chegará a tempo com probabilidade 0.04"
    - "A_{45} chegará a tempo com probabilidade 0.85"
    - "A_{60} chegará a tempo com probabilidade 0.95"

Probabilidade

- A probabilidade proporciona um meio para resumir a incerteza que vem de:
  - Preguiça = falha em enumerar todas as possíveis exceções à regra
  - Ignorância = falta de conhecimento sobre fatos relevantes, condições iniciais

Probabilidade

- Probabilidade subjetiva ou bayesiana
  - Estabelece o estado de crença do agente em uma sentenças, dadas as evidências.
  - Muda quando novas evidências chegam
    - $P(A_{25}|\text{nenhum acidente}) = 0.06$
    - $P(A_{25}|\text{nenhum acidente, 5 a.m.}) = 0.15$

- As sentenças são verdadeiras ou falsas.
  - O que muda é o grau de crença do agente na sentença.
  - Atribuir probabilidade 0 a uma sentença significa acreditar que ela é falsa com certeza absoluta.
  - Atribuir probabilidade 1 a uma sentença significa acreditar que ela é verdadeira com certeza absoluta.

Decisões sob incerteza

- Suponha o seguinte conjunto de crenças:

$$P(A_{25} \text{ chega a tempo} \mid \ldots) = 0.04$$
$$P(A_{90} \text{ chega a tempo} \mid \ldots) = 0.70$$
$$P(A_{120} \text{ chega a tempo} \mid \ldots) = 0.95$$
$$P(A_{1440} \text{ chega a tempo} \mid \ldots) = 0.9999$$

- Que ação o agente deve tomar?
  - Depende de suas **preferências** sob perder o vôo versus o tempo esperando no aeroporto.
    - Teoria da utilidade = representação de preferências
    - Teoria da decisão = teoria da probabilidade + teoria da utilidade

Introdução à probabilidade

- Elemento básico: variável aleatória
  - Análogo à lógica proposicional
    - Mundos possíveis são definidos pela atribuição de valores às variáveis.
  - Cada variável aleatória tem um **domínio** que determina seus valores possíveis.
    - Tipos de domínio
      - **Booleano**, ex.: *Cárie* possui valores em <verdadeiro,falso>
      - **Discreto**, ex.: *Clima* possui valores em <ensolarado, chuvoso, nublado, neve>
      - **Contínuo**, ex.: *Temperatura*

Introdução à probabilidade

- Proposições elementares
  - São construídas através da atribuição de valores a variáveis.
  - Ex.: *Clima* = ensolarado, *Cárie* = falso (abreviado como $\neg$*cárie*)
- Proposições complexas
  - São formadas a partir de proposições elementares e conectivos lógicos padrão
  - Ex.: *Clima* = ensolarado v *Cárie* = falso

Introdução à probabilidade

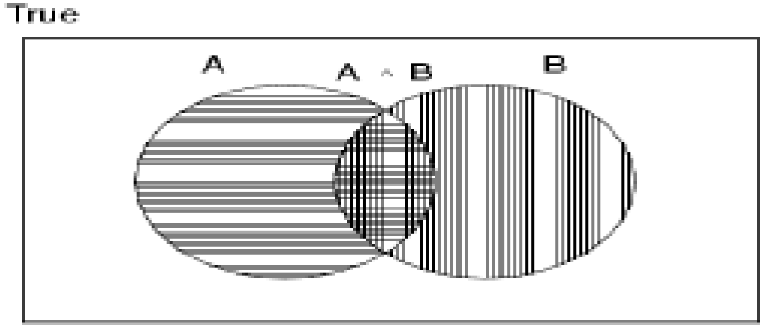
- Evento atômico
  - Especificação completa do estado do mundo sobre o qual o agente está incerto.
    - Uma atribuição de valores a TODAS as variáveis das quais o mundo é formado.
    - Eventos atômicos são mutuamente exclusivos e exaustivos.

Evento atômico: exemplo

- Se o mundo consistir somente de 2 variáveis booleanas (*Cárie* e *DorDeDente*), então há 4 eventos atômicos distintos:
  - *Cárie* = verdadeiro ∧ *DorDeDente* = verdadeiro
  - *Cárie* = verdadeiro ∧ *DorDeDente* = falso
  - *Cárie* = falso ∧ *DorDeDente* = verdadeiro
  - *Cárie* = falso ∧ *DorDeDente* = falso

Axiomas da Probabilidade

- Para quaisquer proposições *A, B*
  - $0 \leq P(A) \leq 1$
  - P(*verdade*) = 1 e P(*falso*) = 0
    - (proposições neces. verdadeiras -- válidas -- prob=1 e proposições neces. falsas – não satisfatíveis -- prob.=0)
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Probabilidade

- A probabilidade de uma proposição é igual à soma das probabilidades dos eventos atômicos em que ela é válida:

$$P(a) = \Sigma_{ei \in e(a)} P(e_i)$$

- Essa equação permite calcular a probabilidade de qualquer proposição dada uma distribuição conjunta total que especifique todos os eventos atômicos.

Probabilidade incondicional ou "a priori"

- É o grau de crença em uma proposição na ausência de outras informações.
  - Exemplos:
    - P(*Cárie* = verdadeiro) = 0.1
    - *P(Clima* = ensolarado) = 0.72
- Distribuição de probabilidades
  - Dá probabilidades a todos os valores possíveis de uma variável aleatória.

P(*Clima*) = <0.72,0.1,0.08,0.1> (normalizado, i.e., soma da 1)

Distribuição de Probabilidade Conjunta

- Probabilidades de todas as combinações de valores de um conjunto de variáveis aleatórias.

P(Clima,*Carie*) = tabela 4 × 2 de valores:

Clima =	ensolarado	chuvoso	nublado	neve
Cárie = verdadeiro	0.144	0.02	0.016	0.02
Cárie = falso	0.576	0.08	0.064	0.08

- Uma distribuição conjunta total especifica a probabilidade de qualquer evento atômico.
  - Qualquer probabilidade nesse domínio pode ser calculada a partir da distribuição conjunta total.

Probabilidade condicional ou "a posteriori"

- É o grau de crença em uma proposição dada a presença de evidências (valores de variáveis aleatórias conhecidos).
  - Exemplos:
    - P(*Cárie* = verdadeiro | *DorDeDente* = verdadeiro) = 0.8
    - P(*Cárie* = verdadeiro | *DorDeDente* = verdadeiro, *Cárie* = verdadeiro) = 1
    - P(*Cárie* = verdadeiro | *DorDeDente* = verdadeiro, *Ensolarado* = verdadeiro) = P(*Cárie* = verdadeiro | *DorDeDente*) = 0.8

- Distribuição condicional
  - **P**(Y|X) fornece o valor de $P(Y=y_i \mid X=x_i)$ para cada valor de *i* e *j* possíveis.

Probabilidade Condicional

- Pode ser definida em termos de probabilidades a priori:
  $P(a \mid b) = P(a \wedge b) / P(b)$ se $P(b) > 0$

- Regra do produto dá uma definição alternativa:
  $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

- Isso pode ser generalizado para distribuições totais: e.g.
  $\mathbf{P}(Clima, \mathit{Cárie}) = \mathbf{P}(Clima \mid \mathit{Cárie})\, \mathbf{P}(\mathit{Cárie})$
  (que é um conjunto de 4 × 2 equações, não uma multiplicação matricial.)

- Regra da cadeia é obtida a partir de aplicações sucessivas da regra do produto:

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}
$$

Inferência Probabilística

- Inferência probabilística: a computação a partir de evidências observadas de probabilidades posteriores para proposições de consulta.
- Inferência com o uso de distribuições conjuntas totais: base de conhecimento a partir da qual são derivadas respostas para todas as consultas.

Exemplo: Inferência Probabilística

- Suponha um domínio com a seguinte distribuição conjunta total:

	dordedente		*¬dordedente*	
	boticão	*¬boticão*	*boticão*	*¬boticão*
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

- Para qualquer proposição ***a***, P(a) é a soma dos eventos atômicos ***w*** onde ***a*** ocorre: $P(a) = \sum_{w:w|=a} P(w)$

Exemplo: Inferência Probabilística

- Suponha um domínio com a seguinte distribuição conjunta total:

	dordedente		*¬dordedente*	
	boticão	*¬boticão*	*boticão*	*¬boticão*
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

- Para qualquer proposição ***a***, P(a) é a soma dos eventos atômicos ***w*** onde ***a*** ocorre: $P(a) = \sum_{w:w|=a} P(w)$

P(*dordedente*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Exemplo: Inferência Probabilística

- Suponha um domínio com a seguinte distribuição conjunta total:

	dordedente		*¬dordedente*	
	boticão	*¬boticão*	*boticão*	*¬boticão*
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

  - Para qualquer proposição ***a***, P(a) é a soma dos eventos atômicos ***w*** onde ***a*** ocorre: $P(a) = \sum_{w:w|=a} P(w)$

P(*dordedente* ∨ *cárie*) = 0.108 + 0.012 + 0.016 + 0.064 + 0.072 + 0.008 = 0.28

Exemplo: Inferência Probabilística

- Podemos calcular probabilidades condicionais:

	dordedente		*¬dordedente*	
	boticão	*¬boticão*	*boticão*	*¬boticão*
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

$$P(\neg cárie | dordedente) = \frac{P(\neg cárie \wedge dordedente)}{P(dordedente)}$$

$$= \frac{0.016+0.064}{0.108+0.012+0.016+0.064}$$

$$= 0.4$$

Normalização

	dordedente		*¬dordedente*	
	boticão	*¬boticão*	*boticão*	*¬boticão*
cárie	.108	.012	.072	.008
¬cárie	.016	.064	.144	.576

- O denominador pode ser visto como uma constante de normalização α.

P(Cárie|dordedente) = α **P**(Cárie,dordedente)
= α [**P**(Cárie,dordedente,boticão) + **P**(Cárie,dordedente,¬boticão)]
= α [<0.108,0.016> + <0.012,0.064>]
= α [<0.12,0.08>]
= <0.6,0.4>]

Redes Bayesianas

Cap 14

Seções 1 – 3

Outline

- Sintaxe
- Semântica
- Distribuições parametrizadas

Redes Bayesianas

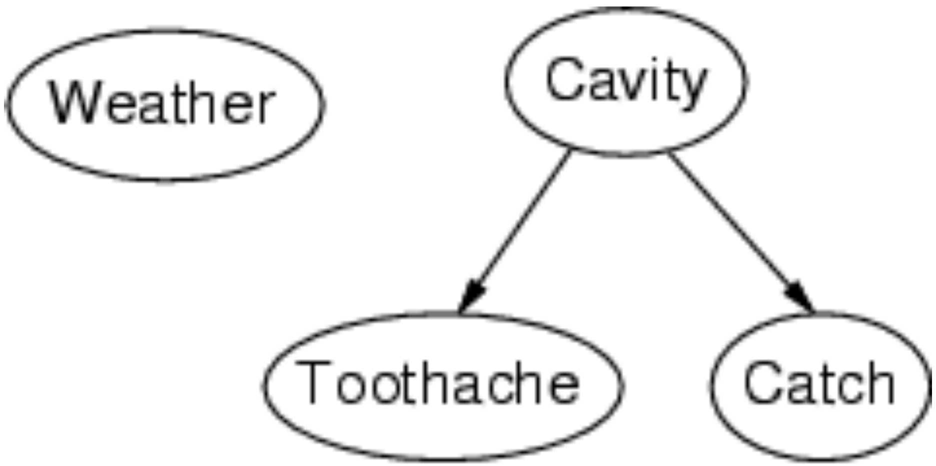
- Estrutura de dados para representar as dependências entre variáveis e fornecer uma especificação concisa de *qualquer* distribuição de probabilidade conjunta total.

- Sintaxe:
  - um conjunto de nós, um para cada variável aleatória
  - grafo direcionado e acíclico (seta = ”influência direta")
  - cada nó tem uma distribuição condicional $\mathbf{P}(X_i \mid \text{Pais}(X_i))$ que quantifica o efeito dos pais sobre o nó

- No caso mais simples, a distribuição condicional é representada como uma tabela de probabilidade condicional (TPC) dada uma distribuição sobre X_i para cada combinação de valores dos pais.

Exemplo

- A topologia de uma rede representa relações de independência condicional :



- *Clima* é independente de outras variáveis
- *Toothache* e *Catch* são condicionalmente independentes dado *Cavity*

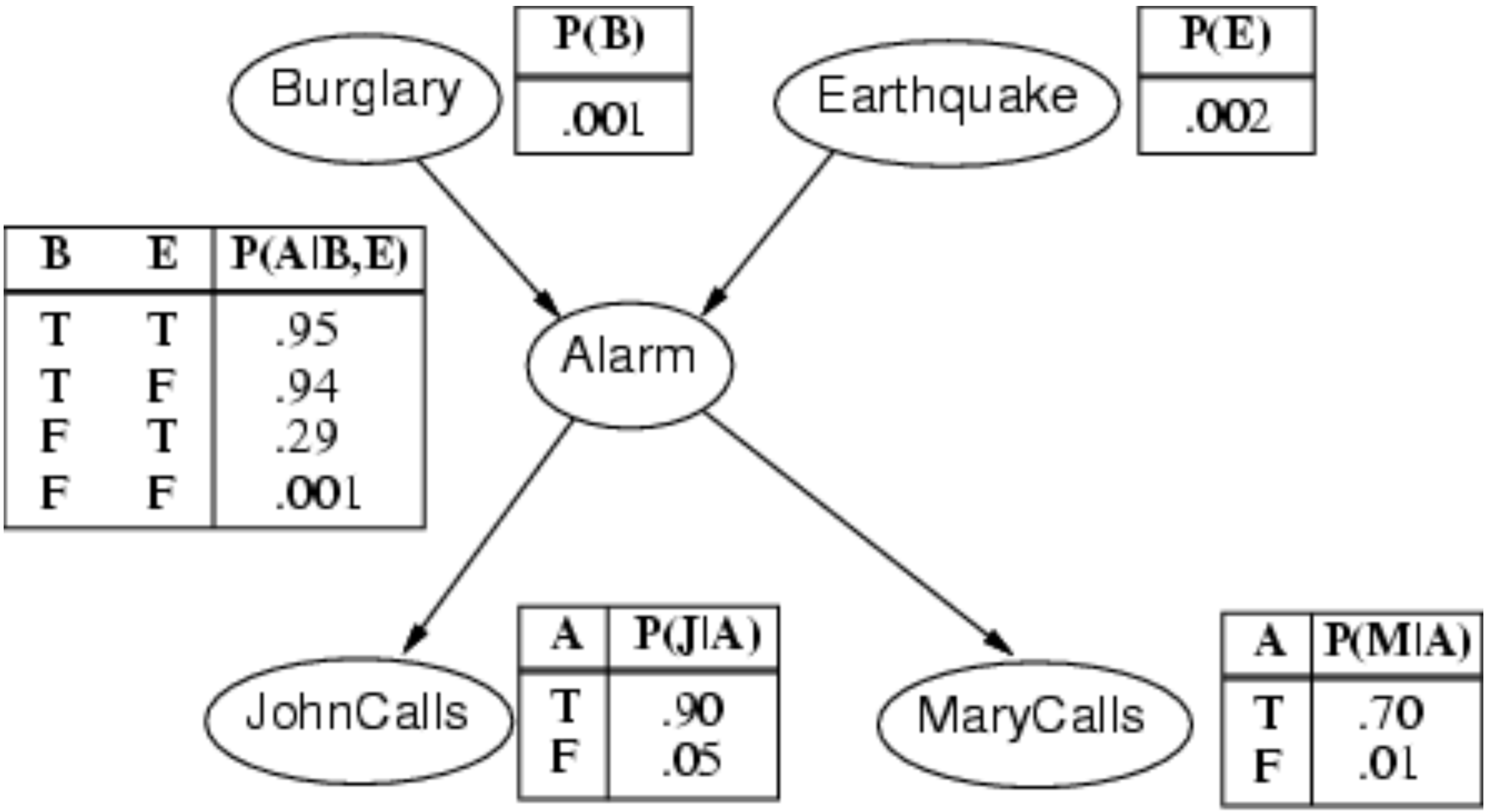
Exemplo

- “ *I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar*? ”

- Variáveis: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- A topologia da rede reflete conhecimento "causal":
  - Um roubo (*burglar) pode ligar o alarme*
  - Um terremoto (*earthquake*) pode ligar o alarme
  - O alarme faz Mary telefonar
  - O alarme faz John telefonar

Exemplo



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Da topologia da rede

- Roubos e terremotos afetam diretamente a probabilidade to alarme tocar;
- Mas o fato de Joao e Maria telefonarem só depende do alarme;
- Desse modo, **a rede representa nossas suposições de que eles não percebem quaisquer roubos diretamente, não notam os terremotos e não verificam antes de ligar!**

As probabilidades...

- ... resumem um conjunto potencialmente infinito de circunstâncias (Maria ouve música alta, João liga qdo toca o telefone; umidade, falta de energia, etc podem interferir no alarme; Joao e maria não estão em casa, etc.
- **Preguiça e ignorância**

Tabelas de probabilidade condicional (TPC)

- Cada linha em uma TPC contém a probabilidade condicional de cada valor de nó para um **caso de condicionamento**;
  - um caso de condicionamento é apenas uma combinação possível de valores para os nós superiores
- Cada linha requer um número *p* para X_i = *true* (prob. para X_i = *false* é *1-p*)
- Um nó sem pais tem apenas uma linha: probabilidade a priori
- Em geral, uma tabela para uma var. booleana com k pais booleanos possui 2^k probabilidades
- Se cada var. não possuir mais do que k pais, a rede completa será $O(n \cdot 2^k)$, para n = número de nós.
  - I.e., cresce linearmente em *n*, vs. $O(2^n)$ da distribuição total

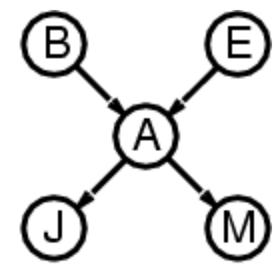
Semântica das RB

- Duas maneiras equivalentes:
  - **Semântica global** (ou numérica): entender as redes como uma representação da distribuição de probabilidade conjunta;
    - indica como construir uma rede
  - **Semântica local** (ou topológica): visualizá-las como uma codificação de uma coleção de declarações de independência condicional.
    - Indica como fazer inferências com uma rede.

Semântica Global

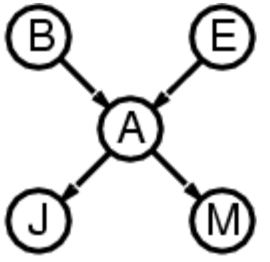
A semântica global (ou numérica) define a distribuição de probabilidade total como o produto das distribuições condicionais locais:

$$\mathbf{P}\,(X_1, \ldots, X_n) = \prod_{i=1} \mathbf{P}\,(X_i \mid Parents(X_i))$$



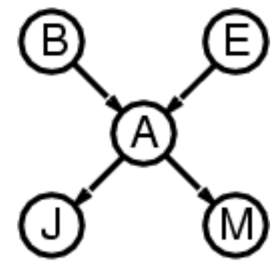
Semântica Global

Se uma rede bayesiana for uma representação da distribuição conjunta, então ela poderá ser utilizada para responder qqr consulta efetuando-se o somatório de todas as entradas conjuntas relevantes



Semântica Global

A semântica global (ou numérica) define a distribuição de probabilidade total como o produto das distribuições condicionais locais:



$$\mathbf{P}\,(X_1, \ldots, X_n) = \prod_{i=1} \mathbf{P}\,(X_i \mid Parents(X_i))$$

e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

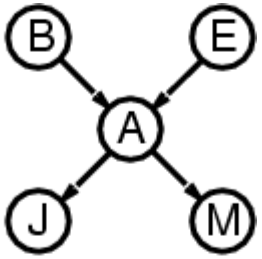
$= \mathbf{P}\,(j \mid a)\, \mathbf{P}\,(m \mid a)\, \mathbf{P}\,(a \mid \neg b, \neg e)\, \mathbf{P}\,(\neg b)\, \mathbf{P}\,(\neg e)$

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$

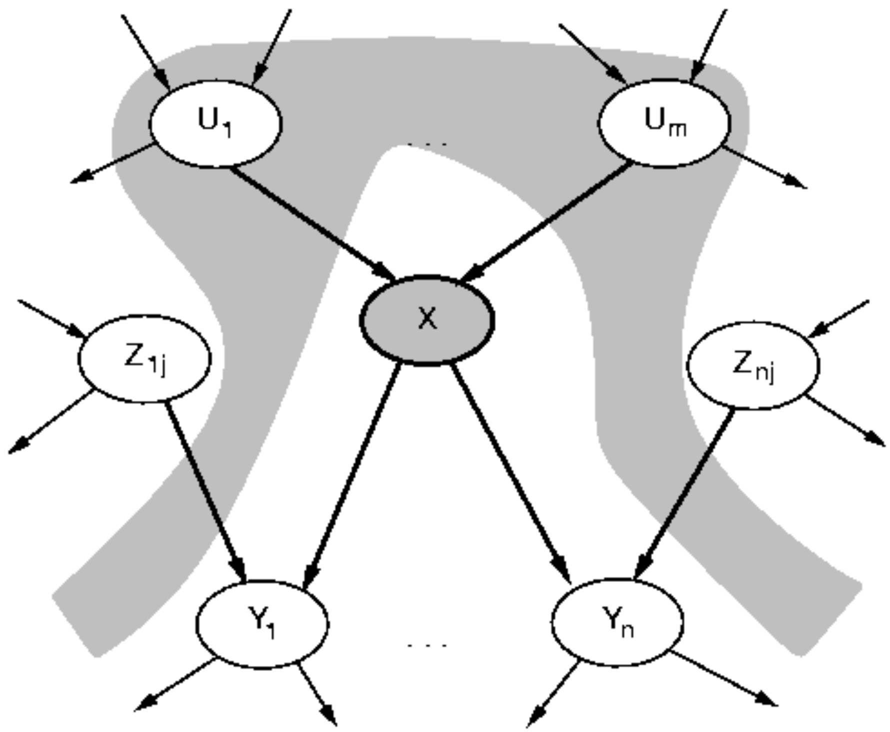
$= 0.00063$

Semântica local

- **Semântica local** (topológica): cada nó é condicionalmente independente de seus não-descendentes, dados seus pais. Ex. **J** é independente de **B** e **E**

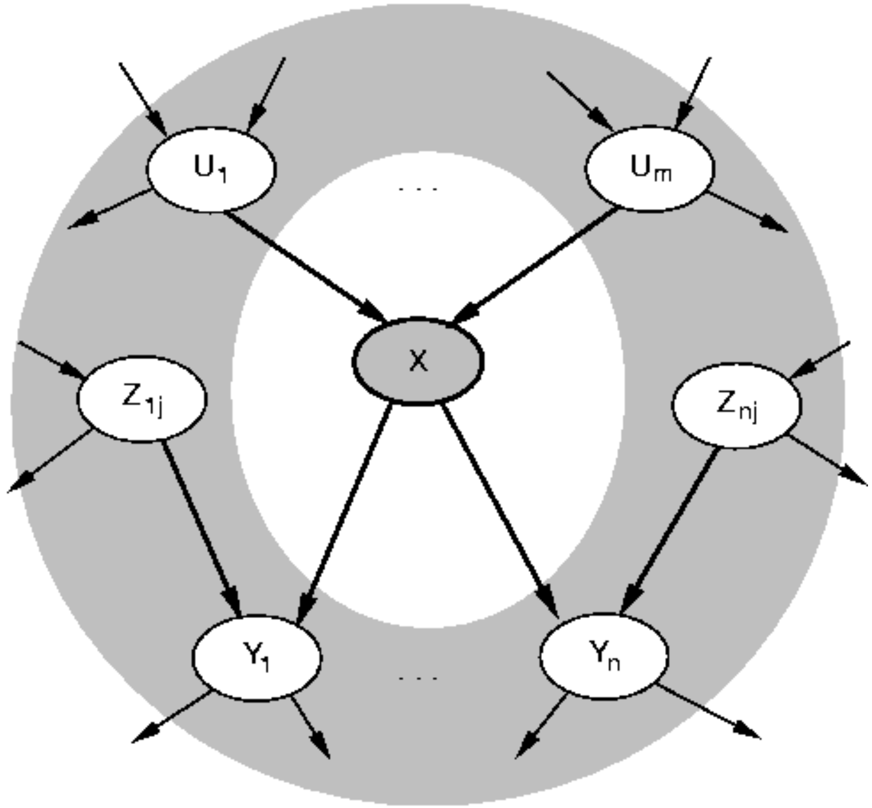


Semântica local



Um nó X é condicionalmente independente de seus não descendentes (ex. Z_{ij}) dados seus pais (U_i)

Semântica local



Um nó X é condicinalmente independente todos os outros dada a sua **cobertura de Markov**

Semântica local e global

- A partir dessas asserções sobre a independência condicional e das TPCs, a distribuição conjunta pode ser reconstruída;
  - desse modo a semântica numérica e topológica são equivalentes.

Construindo uma rede Bayesiana

- 1. Escolher uma ordem para as variáveis aleatórias $X_1, \dots, X_n$
- 2. Para i = 1 à n
  - adicione X_i à rede
  - selecione pais para $X_1, \dots, X_{i-1}$ tais que

$$\mathbf{P}(X_i \mid Pais(X_i)) = \mathbf{P}(X_i \mid X_1, \dots X_{i-1})$$

Esta escolha de pais garante a semântica global:

$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \dots, X_{i-1}) \text{ (regra da cadeia)}$$

$$= \prod_{i=1}^{n} \mathbf{P}(X_i \mid Pais(X_i)) \text{ (por construção)}$$

Ordem para as variáveis

- A ordem correta em que os nós devem ser adicionados consiste em adicionar primeiro as “causas de raiz”, depois as variáveis que elas influenciam e assim por diante, até chegarmos às folhas, que não tem nenhuma influência causal direta sobre as outras variáveis.
- E se escolhermos a ordem “errada”??

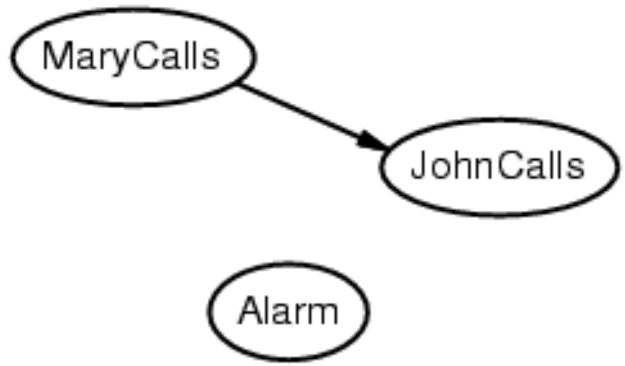
Exemplo

- Assumindo a ordem: *M, J, A, B, E*

$\boldsymbol{P}(J \mid M) = \boldsymbol{P}(J)$?

Exemplo

- Assumindo a ordem: *M, J, A, B, E*

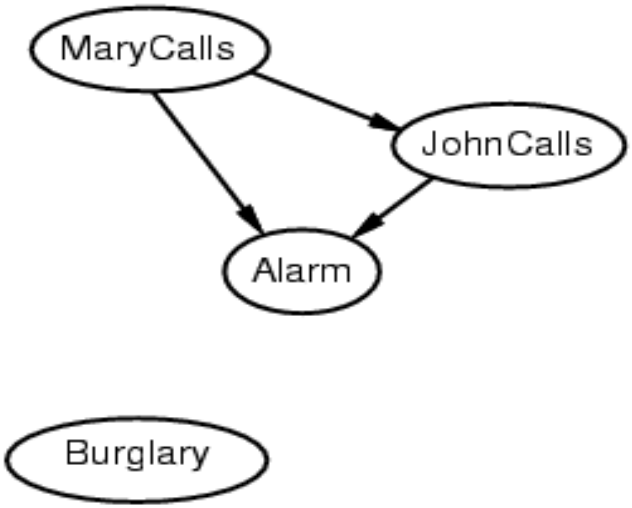


$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

Exemplo

- Assumindo a ordem: *M, J, A, B, E*



$\boldsymbol{P}(J \mid M) = \boldsymbol{P}(J)$? **No**

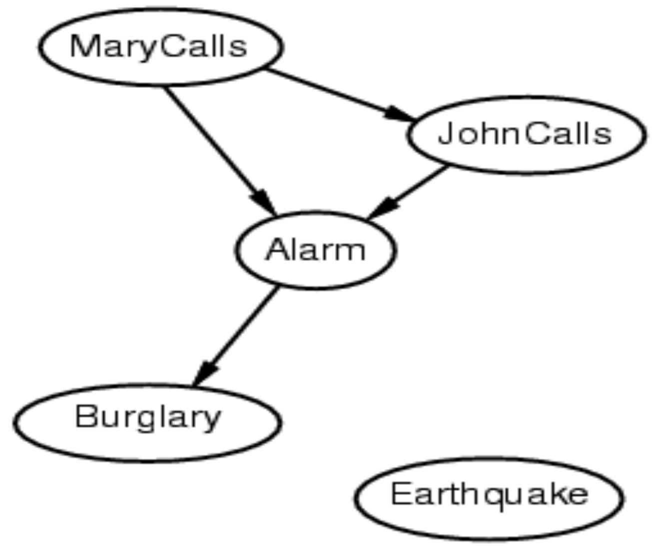
$\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A \mid J)$? $\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A)$? **No**

$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$?

$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B)$?

Exemplo

- Assumindo a ordem: *M, J,*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

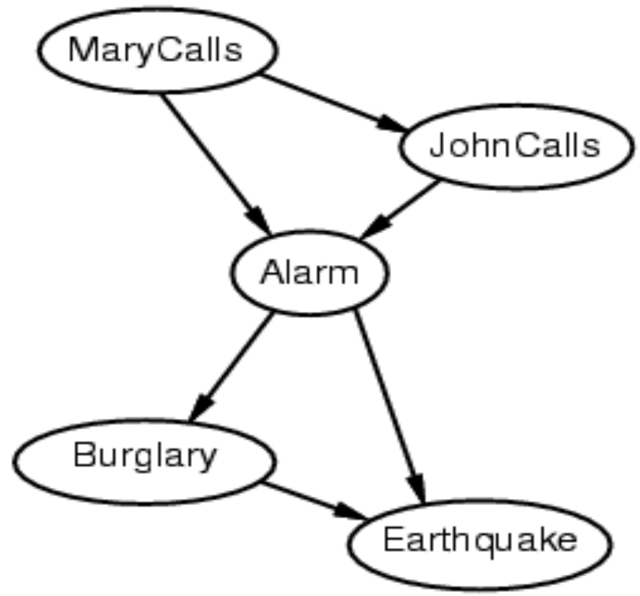
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

$\mathbf{P}(E \mid B, A ,J, M) = \mathbf{P}(E \mid A)$?

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$?

Exemplo



P(*J* | *M*) = ***P***(*J*)? **No**

P(*A* | *J*, *M*) = ***P***(*A* | *J*)? ***P***(*A* | *J*, *M*) = ***P***(*A*)? **No**

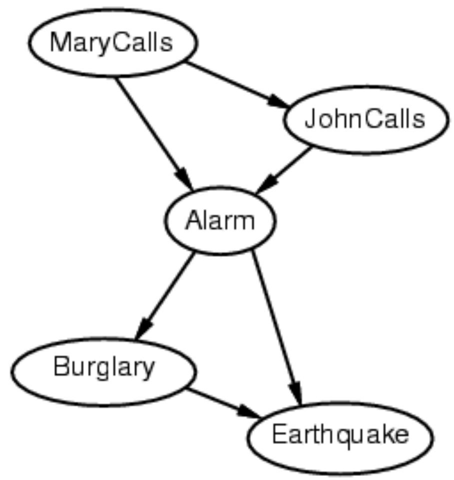
P(*B* | *A*, *J*, *M*) = ***P***(*B* | *A*)? **Yes**

P(*B* | *A*, *J*, *M*) = ***P***(*B*)? **No**

P(*E* | *B*, *A* ,*J*, *M*) = ***P***(*E* | *A*)? **No**

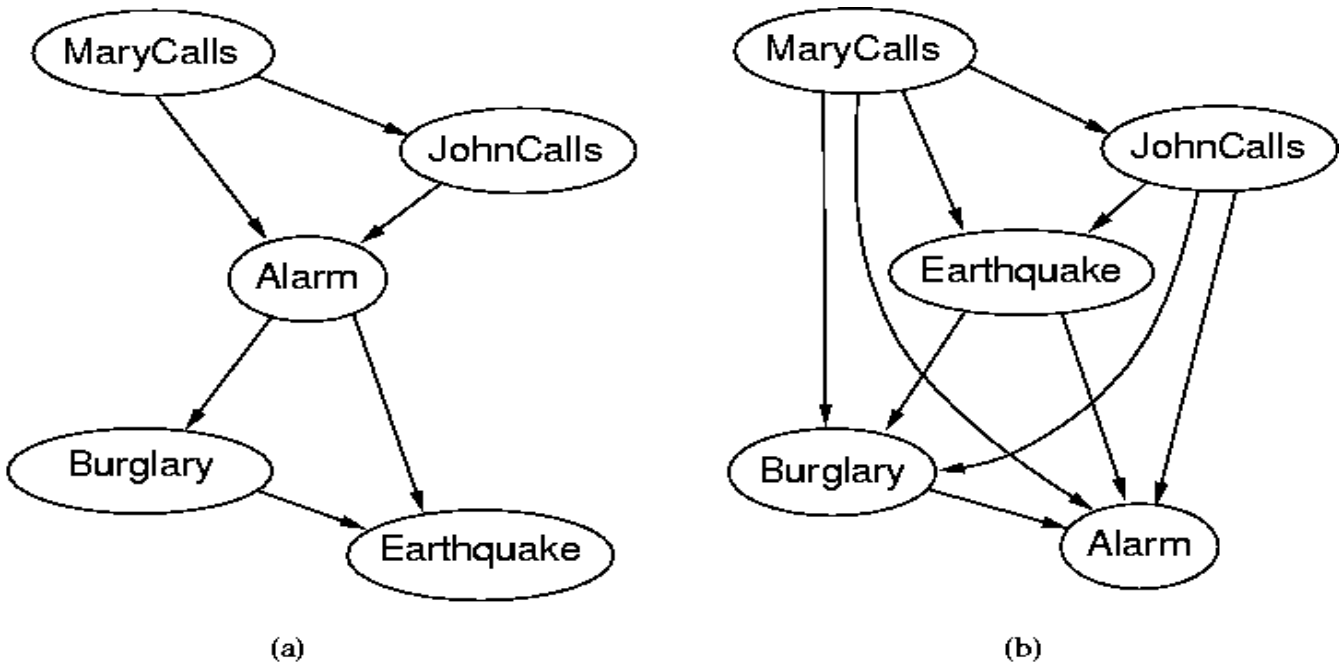
P(*E* | *B*, *A*, *J*, *M*) = ***P***(*E* | *A*, *B*)? **Yes**

Exemplo



- A rede resultante terá dois vínculos a mais que a rede original e exigirá outras probabilidades para serem especificadas
- Alguns dos vínculos apresentam relacionamentos tênues que exigem julgamentos de probabilidade difíceis e antinaturais (prob de *Terremoto*, dados *Roubo* e *Alarme*)
- (Em geral) é melhor pensar de *causas* para *efeitos* (modelo causal) e não do contrário (modelo de diagnóstico)

Exemplo



- Uma ordenação de nós ruim: *MarryCalls*, *JohnCalls*, *Earthquake*, *Burglary* e *Alarm*
- Entretanto, todas as três redes devem representar a mesma distribuição conjunta. As duas últimas só não expressam todas as independências condicionais

Representação eficiente de distribuições condicionais

- Ainda que o número de pais *k* seja reduzido, o preenchimento da TPC para um nó exige até $O(2^k)$ e muita experiência para decidir os casos condicionais.
  - Esse é o pior caso, em que os relacionamentos de pais e filhos é arbitrário
- Em muitos casos podemos utilizar um padrão (***distribuição canônica***) para obter a tabela.

Representação eficiente de distribuições condicionais

- ***Distribuição canônica:***
  - ajustar a distribuição de probabilidades em cada nó a alguma forma padrão;
  - nestes casos a tabela completa pode ser especificada nomeando-se o padrão e fornecendo-se alguns parâmetros.
  - Exemplos:
    - nós determinísticos
    - relacionamentos lógicos ruidosos: *ou-ruidoso*

Representação eficiente: Distribuição canônica

- **Nós determinísticos**: tem seus valores especificados pelos valores de seus pais, sem qualquer incerteza:
  - X = f(Pais(X)) para alguma função f;
  - funções booleanas:
    - Norte Americano $\Leftrightarrow$ Canadense $\vee$US $\vee$ Mexicano
  - relação numérica entre funções contínuas:
    - **pais** *afluentes/escoadouros* **filhos**: Δ*nível da água*
      - Δ*nível da água = afluentes - escoadouros*
    - *Valor mínimo de alguma funçao*

Representação eficiente: Distribuição canônica

- **Ou-ruidoso**
  - $P(\neg fever|\ cold, \neg\ flu, \neg\ malaria) = 0.6$
  - $P(\neg fever|\ \neg\ cold,\ flu, \neg\ malaria) = 0.2$
  - $P(\neg fever|\ \neg\ cold, \neg\ flu,\ malaria) = 0.1$

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	$\mathbf{0.0}$	1.0
F	F	T	0.9	$\mathbf{0.1}$
F	T	F	0.8	$\mathbf{0.2}$
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	$\mathbf{0.6}$
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Representação eficiente: Distribuição canônica

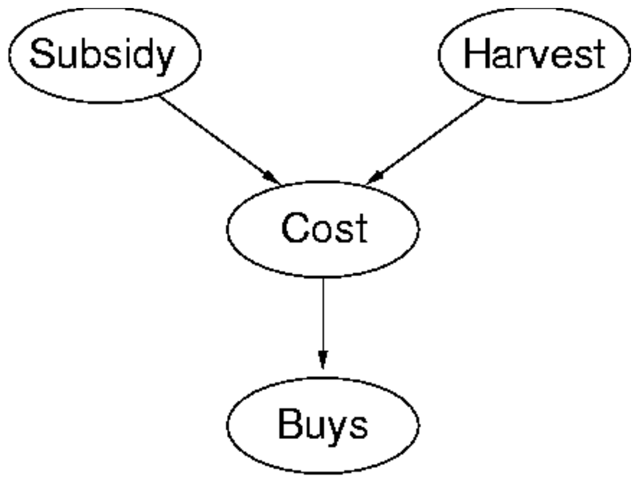
- Ou-ruidoso (em contraste com o ou proposicional)
  - Permite a incerteza sobre a habilidade de cada pai para fazer o filho ser verdadeiro - o relacionamento entre pai e filho pode ser inibido.
    - Todas as causas listadas
    - inibições independentes
    - Assim “febre é falsa sse todos os seus pais verdadeiros são inibidos, e a probabilidade de isso ocorrer é o produto das probabilidades de inibição de cada pai.

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

$P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$
$P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$
$P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$

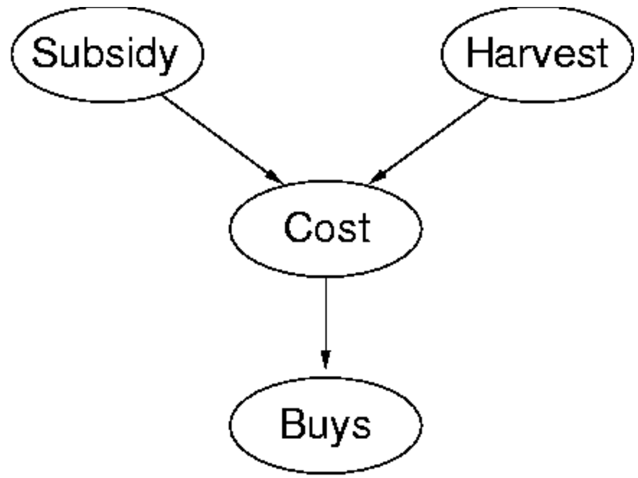
Redes de Bayes Híbridas

- Discretas: *Subsidy?* e *Buys?*



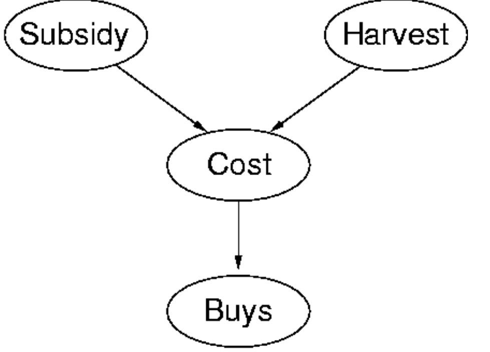
- Dois novos tipos de distr. condicionais:
  - **variável contínua, com pais contínuos e discretos *(Cost)***
  - **Variável discreta com pais contínuos *(Buys?)***

Redes de Bayes Hibridas



- Manipular variáveis contínuas:
  - **Discretização**: repartir os valores possíveis em um conjunto fixo de intervalos
  - **Definir funções de probabilidade padrão** especificadas por um número finito de parâmetros

variável contínua, com pais contínuos e discretos *(Cost)*

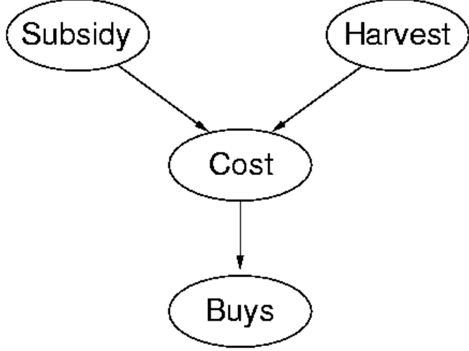


- Para custo: **P**(Custo|Colheita, Subsídio)
  - *O pai discreto (Subsídio) é manipulado por enumeração explícita:*

P(Custo|Colheita, subsídio) e *P(Custo|Colheita, ¬subsídio)*

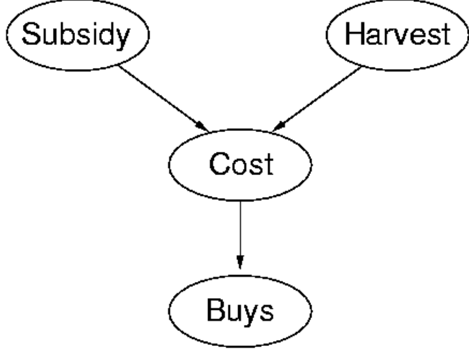
- Para *Colheita* especificamos como a distribuição sobre o custo **c** depende do valor contínuo **h** de colheita.
  - I.e., os parâmetros da distribuição de custo como função de **h**
  - em geral: **distribuição Gaussiana linear**

variável contínua, com pais contínuos e discretos *(Cost)*



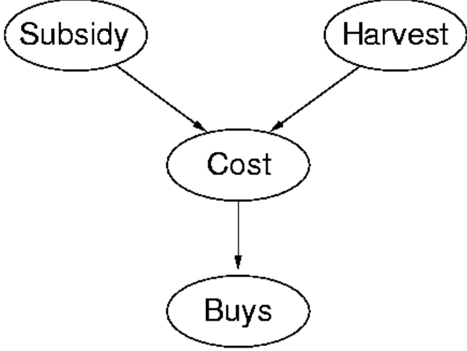
- **distribuição gaussiana linear:** o filho (*Cost*) tem uma distribuição gaussiana cuja média varia linearmente com o valor do pai (*Harvest*), e cujo desvio padrão é fixo:

$$
\begin{aligned}
& P(Cost = c \mid Harvest = h, Subsidy? = true) \\
&= N(a_t h + b_t, \sigma_t)(c) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c-(a_t h + b_t)}{\sigma_t}\right)^2\right)
\end{aligned}
$$



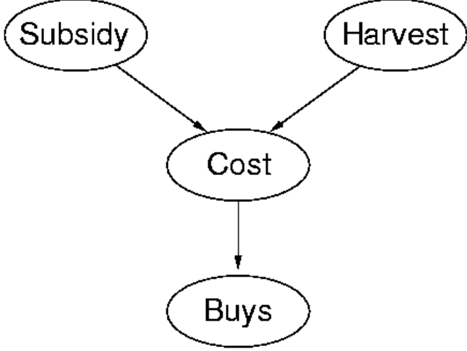
- **distribuição gaussiana linear:** o filho (*Cost*) tem uma distribuição gaussiana cuja média varia linearmente com o valor do pai (*Harvest*), e cujo desvio padrão é fixo:

$$
\begin{aligned}
& P(Cost = c \mid Harvest = h, Subsidy? = true) \\
&= N(a_t h + b_t, \sigma_t)(c) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2} \left(\frac{c - (a_t h + b_t)}{\sigma_t} \right)^2 \right)
\end{aligned}
$$

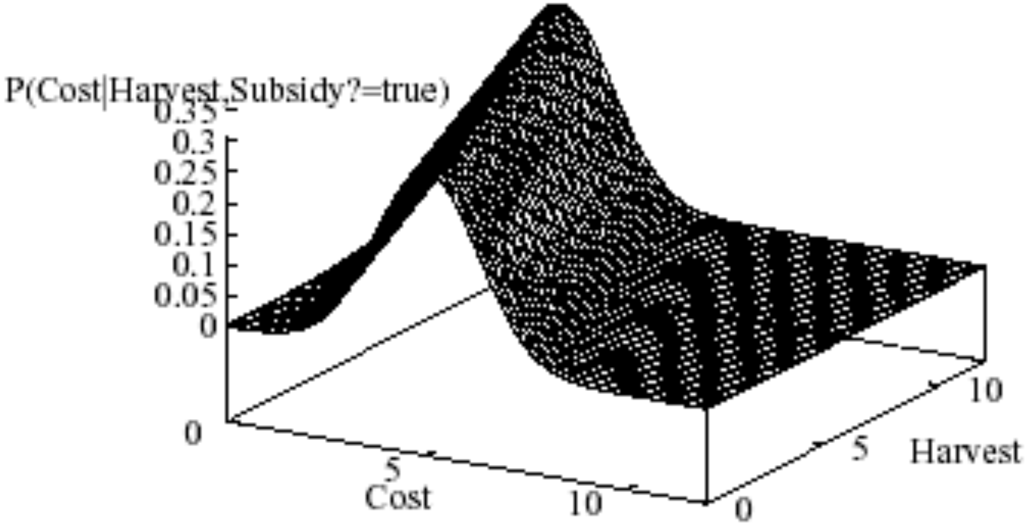


- **distribuição gaussiana linear:** o filho (*Cost*) tem uma distribuição gaussiana cuja média varia linearmente com o valor do pai (*Harvest*), e cujo desvio padrão é fixo:

$$
\begin{aligned}
& P(Cost = c \mid Harvest = h, Subsidy? = true) \\
&= N(a_t h + b_t, \sigma_t)(c) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2} \left(\frac{c - (a_t h + b_t)}{\sigma_t} \right)^2 \right)
\end{aligned}
$$



$$
\begin{aligned}
& P(Cost = c \mid Harvest = h, Subsidy? = true) \\
&= N(a_t h + b_t, \sigma_t)(c) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c-(a_t h + b_t)}{\sigma_t}\right)^2\right)
\end{aligned}
$$



A inclinação é negativa, pq o preço diminui à medida que a quantidade oferecida aumenta

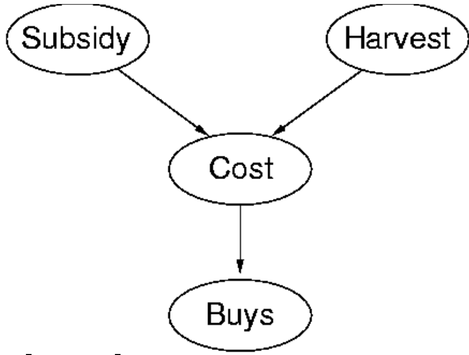
Distribuição gaussiana condicional linear

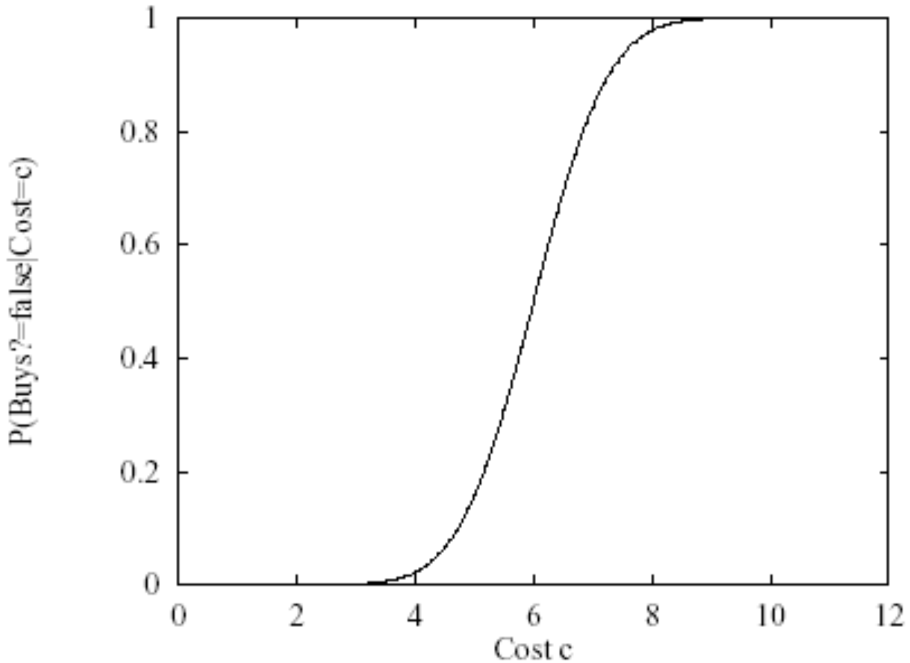
- Uma rede que contém apenas variáveis contínuas com distribuições Gaussianas lineares tem uma distribuição conjunta que é uma **distribuição multivariada** sobre todas as variáveis.
  - superfície em mais de uma dimensão que tem um pico na média (em n dimensões) e decresce para todos os lados.
- Com variáveis discretas (se nenhuma destas é filha de uma var. contínua), a rede define uma **distribuição gaussiana condicional**
  - dada qqr. atribuição às var. discretas, a distribuição sobre as var. contínuas é uma **distribuição gaussiana multivariada.**

variáveis discretas com pais contínuos

- Ex. *Compras*:
  - Podemos supor que o cliente comprará se o preço for baixo e não comprará se for alto e que:
  - A probabilidade de compra varia suavemente em alguma região intermediária
    - A distribuição condicional é semelhante a uma função de limiar “suave” (*soft threshold*)
    - Distribuição **probit** é uma possibilidade...

v.discretas, pais contínuos

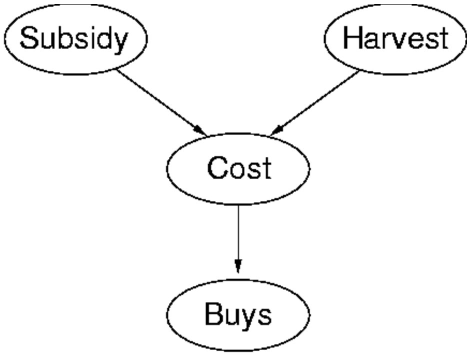


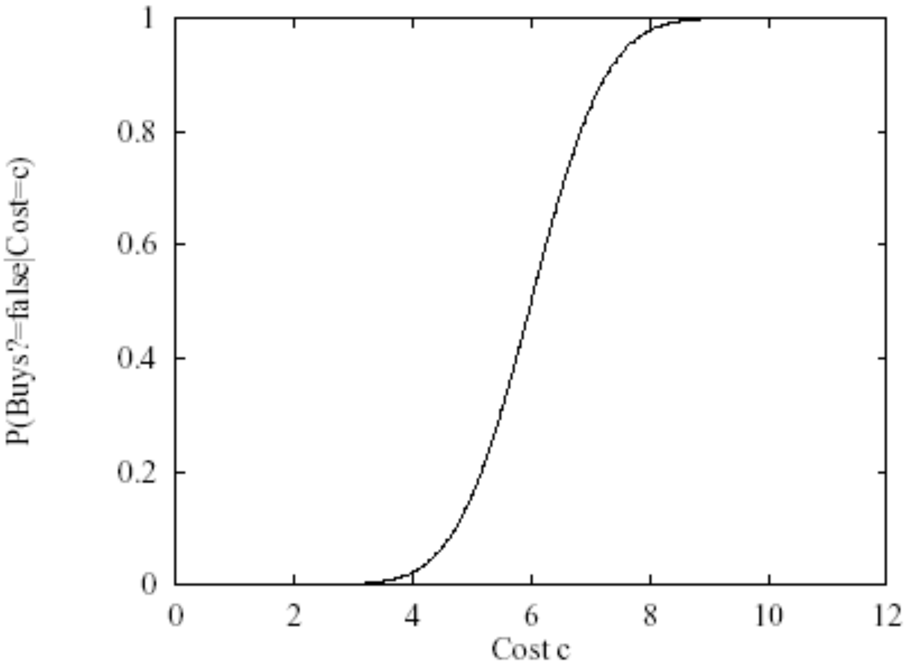


- Probabilidade de *Compra (buys)* dado *Custo (Cost)*: ***limiar suave***
- Distribuição **Probit**:
  - integral da Gaussiana
  - limiar difícil, mas a sua posição precisa está sugeita a ruído gaussiano aleatório

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$$
$$P(Buys?\!=\!true \mid Cost\!=\!c) = \Phi((-c+\mu)/\sigma)$$

v.discretas, pais contínuos





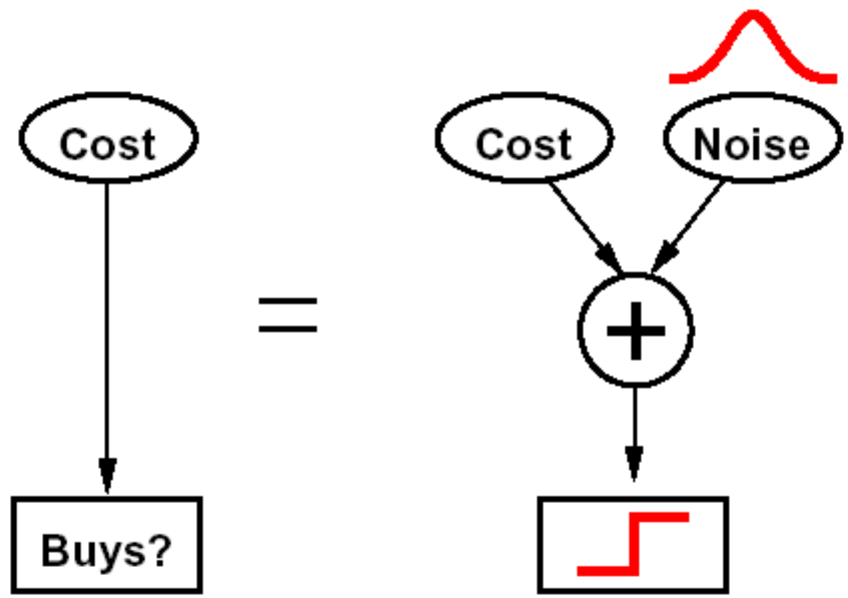
- Probabilidade de *Compra (buys)* dado *Custo (Cost)*: ***limiar suave***
  - O limiar do curto ocorre em torno de , a largura do limiar é a a prob. de compra diminui à medida que o custo aumenta.

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$$

$$P(Buys? = true \mid Cost = c) = \Phi((-c + \mu)/\sigma)$$

Por que **Probit**?

- Possui mais ou menos o formato desejado
- Pode ser visto como um degrau em que a posição é ruidosa



Resumo

- Redes Bayesianas são representações explícitas de independência condicional
- Topologia + TPCs = representações compactas de distribuições conjuntas totais
- Ferramentas poderosas para construir uma representação de um domínio que envolva incerteza.