Rede de Crenças e Lógica Fuzzy

Capítulo 15a

Prof. Dr. Tiago Araújo

INSTITUTO FEDERAL
Pará

Independência

Duas variáveis aleatórias *A B* são (absolutamente) independentes se

$P(A|B) = P(A)$

ou $P(A, B) = P(A|B)P(B) = P(A)P(B)$

por exemplo, A e B são duas moedas moedas atiradas

Se *n* variáveis Booleanas são independentes, o conjunto completo é

$$\mathbf{P}(X_1, \ldots, X_n) = \Pi_i \mathbf{P}(X_i)$$

Portanto, pode ser especificado por apenas *n* número

A independência absoluta é uma exigência muito forte, raramente atendida

Independência Condicional

Considerando o problema do dentista com três variáveis aleatórias:
Toothache, Cavity (cárie), Catch (sonda de aço prende em meu dente)
A distribuição conjunta completa tem $2^3 - 1 = 7$ entradas independentes.

Se eu tenho uma Cavity (cárie), a probabilidade de a sonda prender dentro não depende do clima eu tenho Toothache (uma dor de dente):

$$(1)\ P(Catch|Toothache, Cavity) = P(Catch|Cavity)$$

Por exemplo, *Catch* é condicionalmente independente de *Toothache* dado *Cavity:*
A mesma independência se mantem se eu não tiver uma cárie:

$$(2)\ P(Catch|Toothache, \neg Cavity) = P(Catch|\neg Cavity)$$

Independência Condicional

Declarações equivalentes a (1)

(1a) $P(Toothache|Catch, Cavity) = P(Toothache|Cavity)$ Por quê??

(1b) $P(Toothache, Catch|Cavity) = P(Toothache|Cavity)P(Catch|Cavity)$

Por quê??

A distribuição conjunta completa pode agora ser escrita como

$$\mathbf{P}(Toothache, Catch, Cavity) = \mathbf{P}(Toothache, Catch|Cavity)\mathbf{P}(Cavity)$$
$$= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)$$

Por exemplo, $2 + 2 + 1 = 5$ números independentes (equações 1 e 2 remove 2)

Independência Condicional

Declarações equivalentes a 1

$$P(Toothache|Catch, Cavity)$$
$$= P(Catch|Toothache, Cavity)P(Toothache|Cavity)/P(Catch|Cavity)$$
$$= P(Catch|Cavity)P(Toothache|Cavity)/P(Catch|Cavity)$$ **(de 1)**
$$= P(Toothache|Cavity)$$

(1b) $P(Toothache, Catch|Cavity) = P(Toothache|Cavity)P(Catch|Cavity)$

Por quê?

$$P(Toothache, Catch|Cavity)$$
$$= P(Toothache|Catch, Cavity)P(Catch|Cavity) \text{ (product rule)}$$
$$= P(Toothache|Cavity)P(Catch|Cavity)$$ **(de 1a)**

Redes de Crenças

Uma notação simples e gráfica para afirmações de independência condicional e, portanto, para a especificação compacta de distribuições conjuntas completas

Sintaxe:

- um conjunto de nós, um por variável
- um gráfico dirigido, acíclico (link ≈ "influências diretas")
- uma distribuição condicional para cada nó dado seus pais:

$$\mathbf{P}(X_i | Parents(X_i))$$

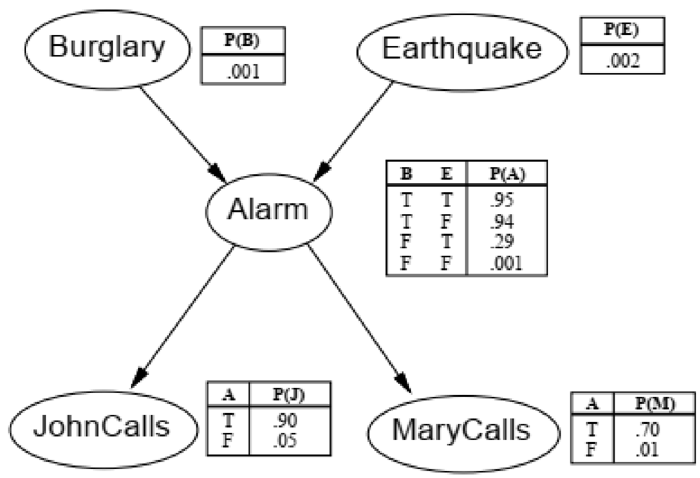
No caso mais simples, a distribuição condicional ressentiu-se como uma tabela de probabilidade condicional (CPT).

Exemplo

Estou no trabalho, o vizinho John liga para dizer que meu alarme está tocando, mas a vizinha Mary não liga. Às vezes é desencadeado por pequenos terremotos. Há um assaltante?

Variáveis: *Burlar, Earthquake, Alarm, JohnCalls, MaryCalls*

A topologia da rede reflete o conhecimento "causal":



Nota: $\leq k.parents \implies O\left(d^k n\right) numbers\ \ vs\ \ O(d^n)$

Semânticas

A semântica "global" define a distribuição conjunta completa como o produto da distribuição condicional local:

$$\mathbf{P}(X_1, \ldots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

Por exemplo,

$$\mathbf{P}(X_1, \ldots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

e.g., $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$ **É dada por??**

$=$

Semânticas

A semântica "global" define a distribuição conjunta completa como o produto da distribuição condicional local

$$\mathbf{P}(X_1, \ldots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

Por exemplo,

$$\mathbf{P}(X_1, \ldots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

e.g., $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$ **É dada por??**

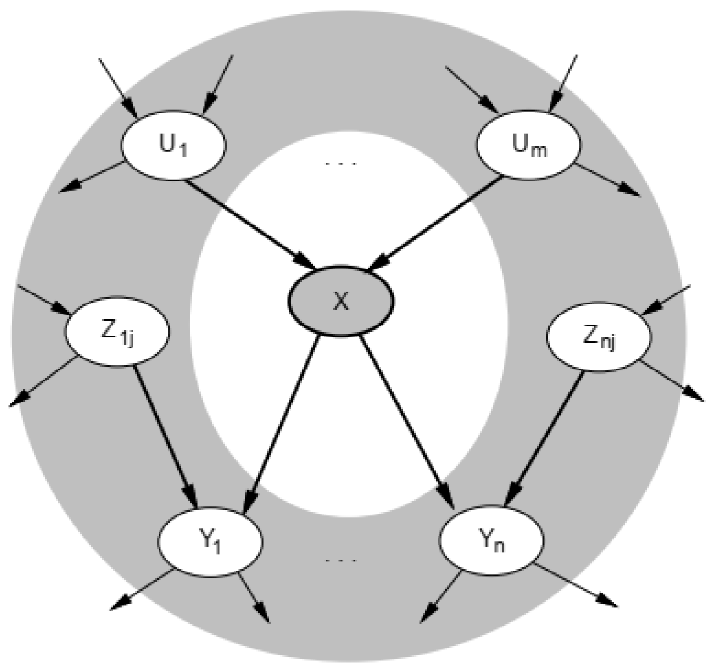
$$= P(\neg B) P(\neg E) P(A | \neg B \wedge \neg E) P(J | A) P(M | A)$$

Semântica "local": cada nó é condicionalmente independente de seus não-descendentes, dado que seus pais

Teorema: Semântica local $\Leftrightarrow$ semântica global

Manta de Markov

Cada nó é condicionalmente independente de todos os outros, dada a Manta de Markov: parents + children + children's parents



Construindo a rede de crenças

Precisa de um método tal que uma série de afirmações locais de independência condicional garanta a necessária semântica global.

1. Escolha uma ordenação de variáveis *X1, ..., Xn*
2. Para *i = 1 até n*

   adicione *Xi* na rede

   $$= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

   selecione parents ue

   $$\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$$

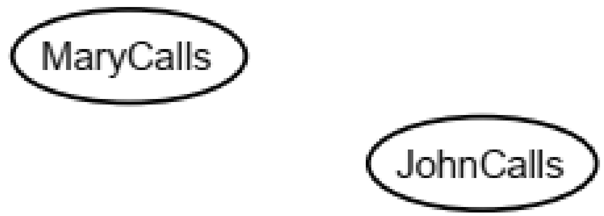
Essa escolha de pais garante a semântica global:

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

$$= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \quad \text{por construção}$$

Exemplo

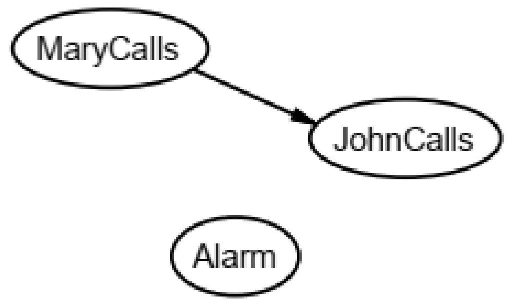
Suponha que nós escolhemos ordenar *M, J, A, B, E*



$P(J|M) = P(J)$?

Exemplo

Suponha que nós escolhemos ordenar *M, J, A, B, E*

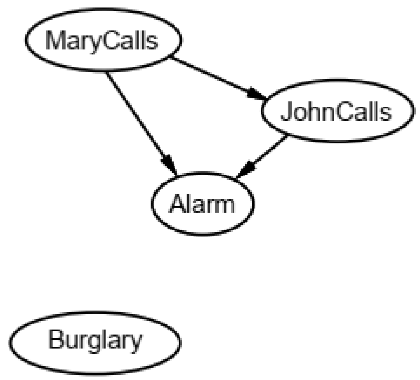


$P(J|M) = P(J)?$ *Não*

$P(A|J,M) = P(A|J)?$ $P(A|J,M) = P(A)?$

Exemplo

Suponha que nós escolhemos ordenar *M, J, A, B, E*



$P(J|M) = P(J)?$ *Não*

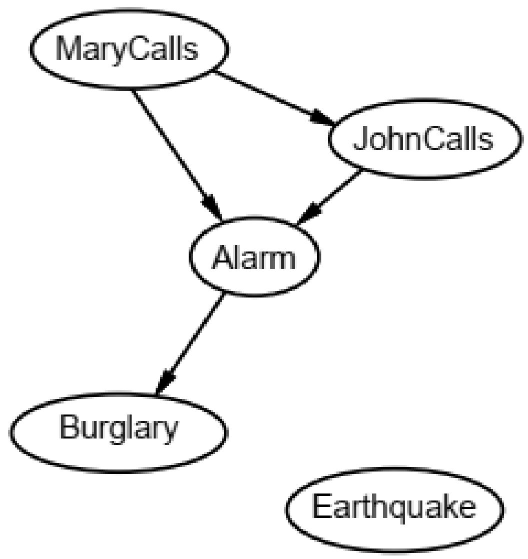
$P(A|J,M) = P(A|J)?$ $P(A|J,M) = P(A)?$ *Não*

$P(B|A,J,M) = P(B|A)?$

$P(B|A,J,M) = P(B)?$

Exemplo

Suponha que nós escolhemos ordenar *M, J, A, B, E*



$P(J|M) = P(J)$? *Não*

$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? *Não*

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

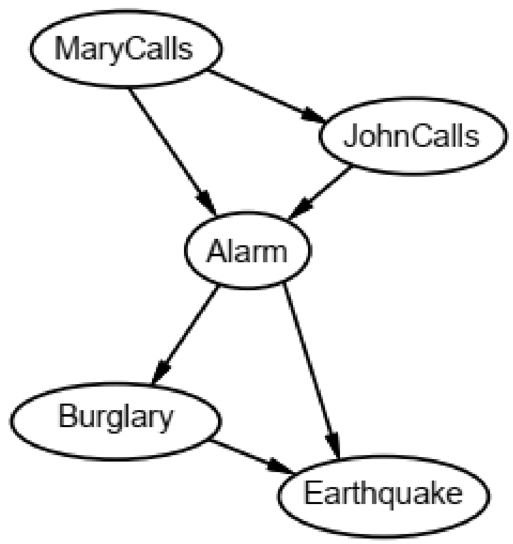
$P(B|A,J,M) = P(B)$? *Não*

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

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

Exemplo

Suponha que nós escolhemos ordenar *M, J, A, B, E*



$P(J|M) = P(J)$? *Não*

$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? *Não*

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

$P(B|A,J,M) = P(B)$? *Não*

$P(E|B,A,J,M) = P(E|A)$? *Não*

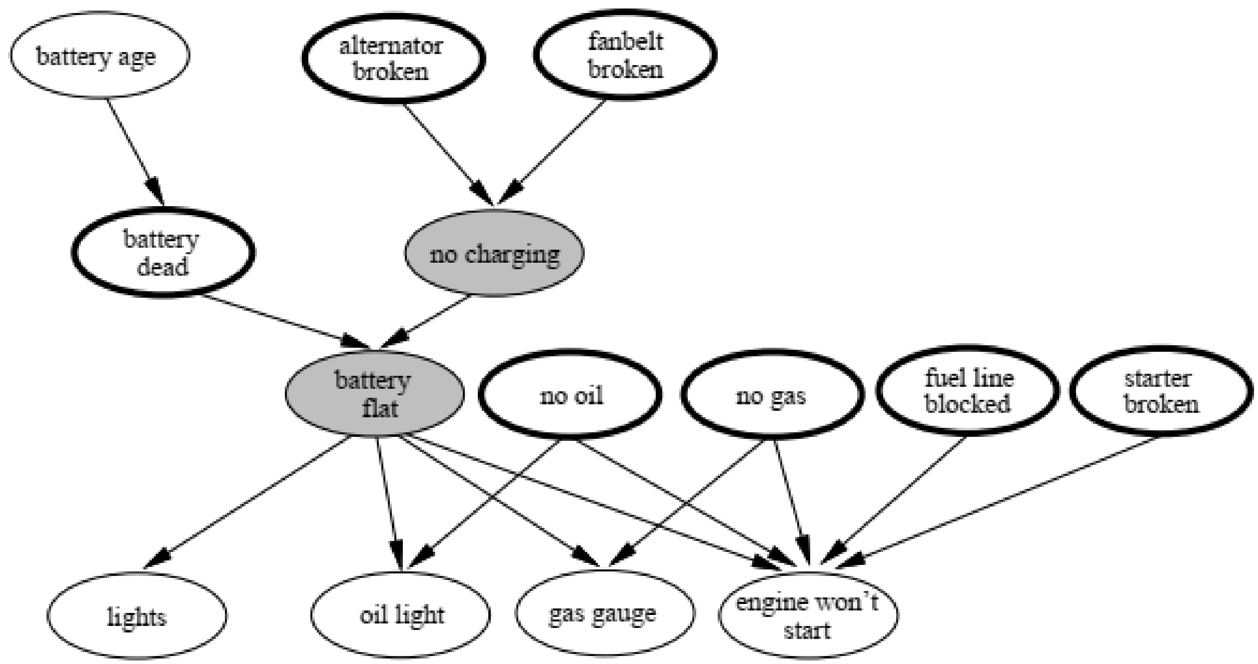
$P(E|B,A,J,M) = P(E|A,B)$? Sim

Exemplo: Diagnóstico de Carros

Evidência inicial: o motor não pega

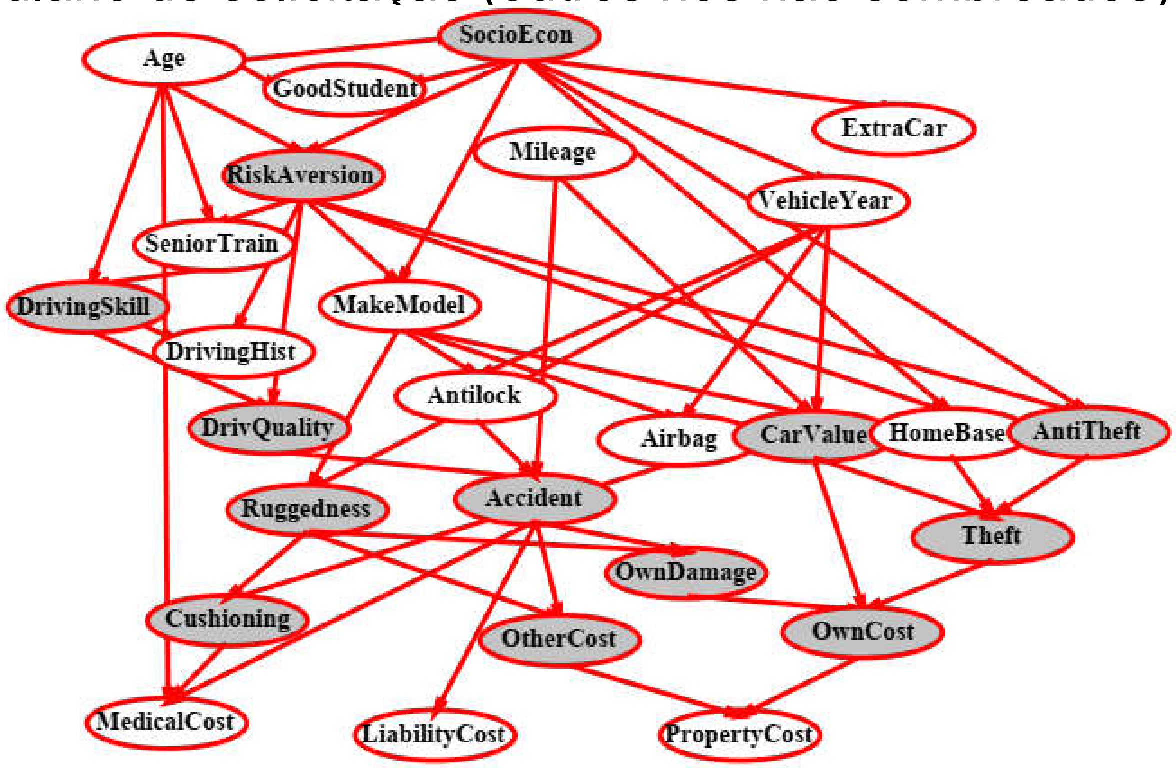
Variáveis testáveis (ovais finos), variáveis de diagnóstico (ovais grossos)

Variáveis ocultas (sombreadas) garantem estrutura esparsa, reduzem parâmetros



Exemplo: Seguro do carro

Prever custos de reclamação (médicos, responsabilidade civil, propriedade) com dados no formulário de solicitação (outros nós não sombreados)



Distribuições condicionais compactas

CPT cresce exponencialmente com o no.de pais CPT torna-se infinito com pais ou filhos de valor contínuo

Solução: distribuições canônicas que são definidas de forma compacta

Os nós determinísticos são o caso mais simples:

$X = f(Parents(X))$ para algumas funções *f*

Por exemplo: Funções Booleanas

$$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$$

Por exemplo, relações numerais entre variáveis contínuas

$$\frac{\partial Level}{\partial t} = \text{inflow} + \text{precipation - outflow - evaporation}$$

Distribuições condicionais compactas

Noisy-OR distribuições modelo de múltiplas causas não-interativas

1) Pais $U_1 \ldots U_k$ incluindo todas as causas (pode adicionar um (nó de fuga

2) Probabilidade de falha independente q_i para cada causa isolada

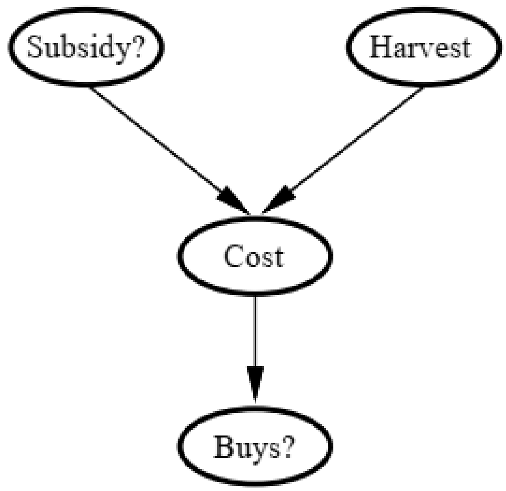
$$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \Pi_{i=1}^{j} q_i$$

$Cold$	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	$0.012 = 0.6 \times 0.2 \times 0.1$

Números de parâmetros *lineares* em número de pais

Redes híbridas (discretas+contínuas)

Dicreto (*subsidy? e Buys?);* contínuo (*Havest e Cost*)



Opção 1: dicretização possivelmente grandes erros, CPT grande
Opção 2: famílias canônicas finamente parametrizadas1) Variável contínua, pais discretos+ contínuos (por exemplo, *Cost)*
2) Variável discreta, pais contínuos (por exemplo, *Buys?*)

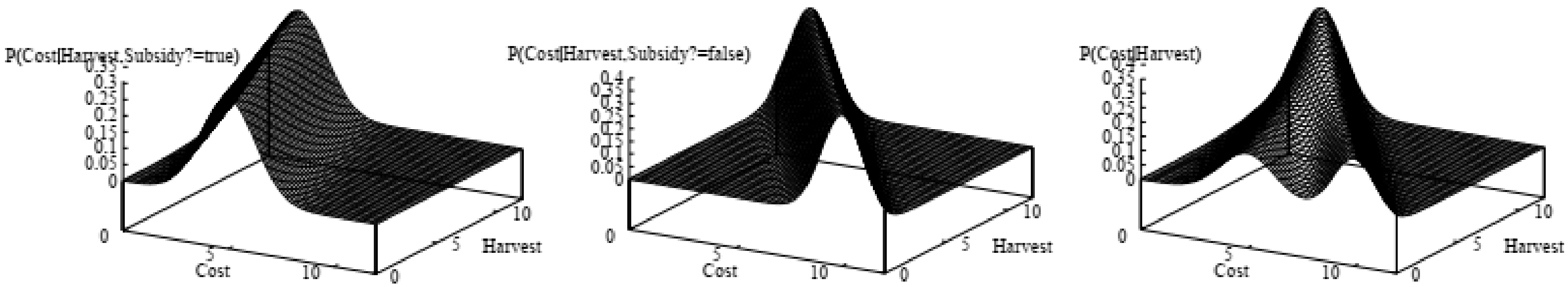
Variáveis Filhas Contínuas

Necessidade de uma função de densidade condicional para a variável filha dada aos pais contínuos, para cada possível atribuição aos pais discretos.
O mais comum é o modelo *linear gaussiano*, por exemplo

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

O *Cost* médio varia linearmente com a *Harvest*, a variação é fixa A variação linear não é razoável em toda a faixa, mas funciona OK se a faixa <u>provável</u> de *Harvest* for estreita.

Variáveis Filhas Continuas



Rede totalmente contínua com distribuições LG

$\implies$ conjunto completo é uma Gaussiana multivariável

A rede LG discreta + continua é uma rede Gaussiana condicional, ou seja; uma Gaussiana multivariada sobre todas as variáveis contínuas para cada combinação de valores discretos de variáveis

Variável discreta com pais contínuos

Probabilidade de *Buy*? dado o *Cost* deve ser um limite "suave":

epsfxsize=0,6
fig{\i1}{probit.ps}}{file{\i1}{graphs}{probit.ps}

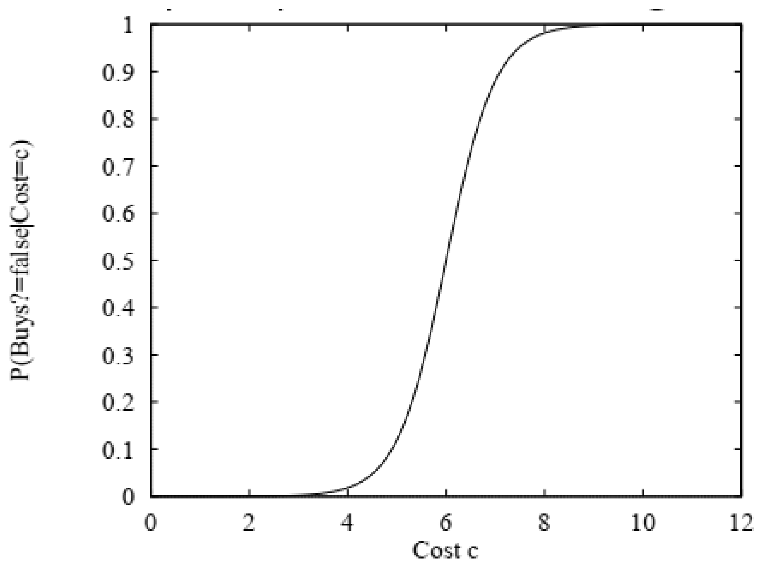
A distribuição Probit usa integral de Gaussian:
  Phi(x) = {^x}{^x} N(0,1)(x) dx
  P(Buys? true given Cost c) = \Phi((-c + \mu)/sigma)$ $(Buys? true given Cost c) = \Phi((-c + \mu)/sigma)$(Buys? true given Cost c)
Pode ver como limiar duro cuja localização está sujeita a ruído

Variáveis Discretas cont.

A distribuição Sigmoid (ou logit) também é utilizada em redes neurais:

$$P(Buys? = true \mid Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$

O Sigmoid tem forma semelhante à sonda, mas com caudas muito mais longas:



Tarefas de Inferência

Consultas simples: calcular marginal posterior $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
Por exemplo $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

Consultas Conjuntivas: $\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$

Decisões ótimas: as redes de decisão incluem informações de utilidade pública; inferência probabilística necessária para
Valores de informação: quais evidências a serem buscadas a seguir?

Análise de sensibilidade: quais valores de probabilidade são mais críticos?
Explicação: por que eu preciso de um novo motor de arranque?

Inferência por enumeração

Uma forma ligeiramente inteligente de extrair variáveis da junta sem construir sua representação explícita

Consulta simples sobre a rede de arrombamento:

$$
\begin{aligned}
&\mathbf{P}(B|J=true, M=true) \\
&= \mathbf{P}(B, J=true, M=true)/P(J=true, M=true) \\
&= \alpha\mathbf{P}(B, J=true, M=true) \\
&= \alpha\Sigma_e\Sigma_a\mathbf{P}(B, e, a, J=true, M=true)
\end{aligned}
$$

Reescrever entradas conjuntas completas usando o produto das entradas do CPT:

$$
\begin{aligned}
&P(B=true|J=true, M=true) \\
&= \alpha\Sigma_e\Sigma_a P(B=true)P(e)P(a|B=true, e)P(J=true|a)P(M=true|a) \\
&= \alpha P(B=true)\Sigma_e P(e)\Sigma_a P(a|B=true, e)P(J=true|a)P(M=true|a)
\end{aligned}
$$

Algoritmo de Enumeração

Exaustiva enumeração em profundidade: $O(n)\ espaço, O\ (d^n)\ tempo$

```
ENUMERATIONASK(X,e,bn) returns a distribution over X
inputs: X, the query variable
        e, evidence specified as an event
        bn, a belief network specifying joint distribution P(X_1,...,X_n)

  Q(x) ← a distribution over X
  for each value x_i of X do
      extend e with value x_i for X
      Q(x_i) ← ENUMERATEALL(VARS[bn],e)
  return NORMALIZE(Q(X))
```

```
ENUMERATEALL(vars,e) returns a real number
  if EMPTY?(vars) then return 1.0
  else do
      Y ← FIRST(vars)
      if Y has value y in e
          then return P(y | Pa(Y)) × ENUMERATEALL(REST(vars),e)
          else return Σ_y P(y | Pa(Y)) × ENUMERATEALL(REST(vars),e_y)
              where e_y is e extended with Y = y
```

Inferência por eliminação de variável

A enumeração é ineficiente: cálculo repetido

Por exemplo cálculo $P(J=true|a)P(M=true|a)$ para cada valor de ***e***

Eliminação de variáveis: realizar totalizações da direita para a esquerda, armazenando resultados intermediários (fatores) para evitar a recomputação

$$
\begin{aligned}
&\mathbf{P}(B|J=true, M=true) \\
&= \alpha \underbrace{\mathbf{P}(B)}_{B} \Sigma_e \underbrace{P(e)}_{E} \Sigma_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(J=true|a)}_{J} \underbrace{P(M=true|a)}_{M} \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) P(J=true|a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a f_A(a,b,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) f_{\bar{A}JM}(b,e) \text{ (sum out } A) \\
&= \alpha \mathbf{P}(B) f_{\bar{E}\bar{A}JM}(b) \text{ (sum out } E) \\
&= \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}
$$

Eliminação de Variável: Operações Básicas

- Produto pontual dos fatores *f1* e *f2:*

$$f_1(x_1,\ldots,x_j,y_1,\ldots,y_k)\times f_2(y_1,\ldots,y_k,z_1,\ldots,z_l)$$
$$= f(x_1,\ldots,x_j,y_1,\ldots,y_k,z_1,\ldots,z_l)$$

Por exemplo, $f_1(a,b)\times f_2(b,c) = f(a,b,c)$

Soma de uma variável a partir de um produto de fatores: mover qualquer fator constante para fora da soma:

$$\Sigma_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \, \Sigma_x \, f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_{\bar{X}}$$

assumindo f1,...., fi não dependem de X

Algoritmo de Eliminação de Variável

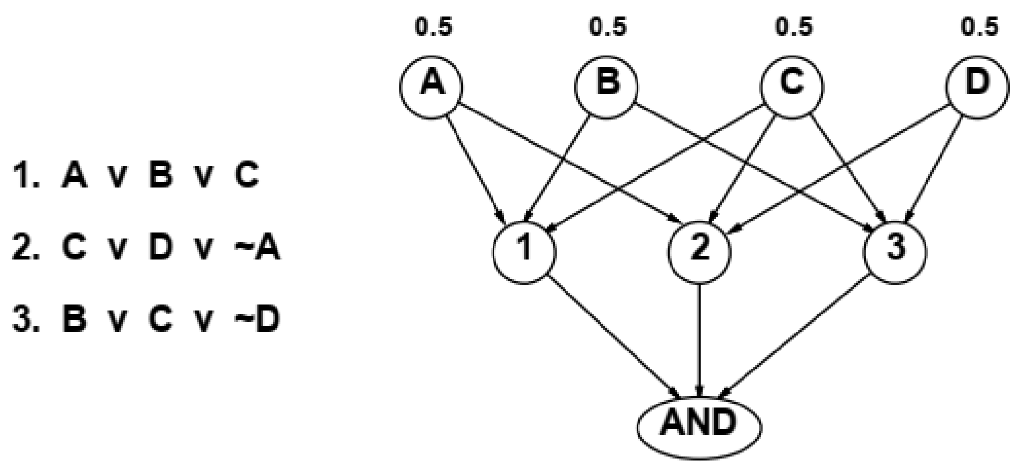
function ELIMINATIONASK(X,$\mathbf{e}$,bn) **returns** a distribution over X
    inputs: X, the query variable
            $\mathbf{e}$, evidence specified as an event
            bn, a belief network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

    if $X \in \mathbf{e}$ **then return** observed point distribution for X
    factors ← [ ]; *vars* ← REVERSE(VARS[bn])
    for each *var* **in** *vars* **do**
        factors ← [MAKEFACTOR(*var*, $\mathbf{e}$)|*factors*]
        if *var* is a hidden variable **then** *factors* ← SUMOUT(*var*,*factors*)
    return NORMALIZE(POINTWISEPRODUCT(*factors*))

Complexidade da inferência exata

Redes conectadas individualmente ou (politrizes) :
quaisquer dois nós são conectados por, no máximo, um caminho (não direcionado)
Os custos de tempo e espaço da eliminação variável são

Multiplicar as redes conectadas:
pode reduzir o 3SAT à inferência exata $\Longrightarrow$ NP-Difícil
equivalente a *contar* os modelos 3SAT $\Longrightarrow$ #P- Completo



Inferência por simulação estocástica

Ideia básica:

1)Tirar *N* amostras de uma distribuição de amostras *S*

2) Calcule uma probabilidade posterior aproximada $\hat{P}$

Mostre que isto converge para a verdadeira probabilidade *P*

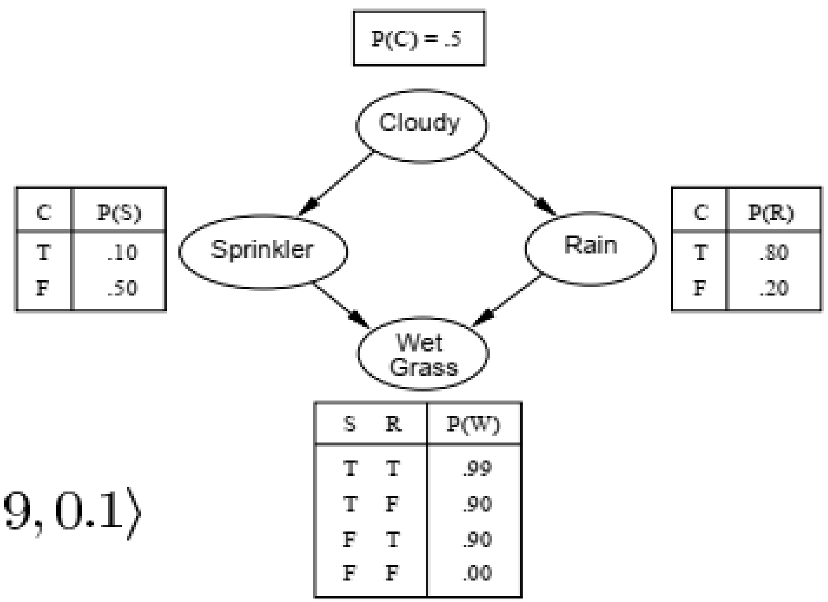
Resumo:

- Amostragem a partir de uma rede vazia
- Rejeição de amostras: rejeitar amostras que discordam das provas
- Ponderação da probabilidade: usar provas para pesar amostras
- MCMC: amostra de um processo estocástico cuja distribuição estacionária insere o verdadeiro posterior

Amostragem a partir de uma rede vazia

function PRIORSAMPLE(bn) **returns** an event sampled from $\mathbf{P}(X_1, \ldots, X_n)$ specified by bn
    $\mathbf{x} \leftarrow$ an event with n elements
    for $i = 1$ **to** n **do**
        $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid Parents(X_i))$
    return x

$\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$
    sample $\rightarrow$ $true$
$\mathbf{P}(Sprinkler|Cloudy) = \langle 0.1, 0.9 \rangle$
    sample $\rightarrow$ $false$
$\mathbf{P}(Rain|Cloudy) = \langle 0.8, 0.2 \rangle$
    sample $\rightarrow$ $true$
$\mathbf{P}(WetGrass|\neg Sprinkler, Rain) = \langle 0.9, 0.1 \rangle$
    sample $\rightarrow$ $true$



Amostragem a partir de uma rede vazia

Probabilidade de que o *PRIORSAMPLE* gere um determinado evento

$$S_{PS}(x_1 \ldots x_n) = \Pi_{i=1}^{n} P(x_i | Parents(X_i)) = P(x_1 \ldots x_n)$$

isto é, então a probabilidade anterior verdadeira

Seja $N_{PS}(\mathbf{Y} = \mathbf{y})$ número de amostras geradas para as quais $\mathbf{Y} = \mathbf{y}$ para qualquer conjunto de variáveis **Y.**

Então $\hat{P}(\mathbf{Y} = \mathbf{y}) = N_{PS}(\mathbf{Y} = \mathbf{y})/N$ e

$$\begin{aligned} \lim_{N \to \infty} \hat{P}(\mathbf{Y} = \mathbf{y}) &= \Sigma_{\mathbf{h}} S_{PS}(\mathbf{Y} = \mathbf{y}, \mathbf{H} = \mathbf{h}) \\ &= \Sigma_{\mathbf{h}} P(\mathbf{Y} = \mathbf{y}, \mathbf{H} = \mathbf{h}) \\ &= P(\mathbf{Y} = \mathbf{y}) \end{aligned}$$

Ou seja, as estimativas derivadas do *PRIORSAMPLE* são consistentes

Reieicão de amostra

$\hat{\mathbf{P}}(X|\mathbf{e})$ estimado a partir de amostra de acordo com o **e**

function REJECTIONSAMPLING(X,**e**,bn,N) **returns** an approximation to $P(X|\mathbf{e})$
  $\mathbf{N}[X] \leftarrow$ a vector of counts over X, initially zero
  for $j = 1$ to N **do**
    $\mathbf{x} \leftarrow$ PRIORSAMPLE(bn)
    if x is consistent with **e then**
      $\mathbf{N}[x] \leftarrow \mathbf{N}[x]+1$ where x is the value of X in $\mathbf{x}$
  return NORMALIZE($\mathbf{N}[X]$)

Por exemplo, estimativa $\mathbf{P}(Rain|Sprinkler = true)$ usando 100 amostras 27 amostras têm $Sprinkler = true$
Desses, 8 tem $Rain = true$ 19 tem $Rain = false$.

$$\hat{\mathbf{P}}(Rain|Sprinkler = true) = \text{NORMALIZE}(\langle 8, 19\rangle) = \langle 0.296, 0.704\rangle$$

Semelhante ao procedimento básico de estimativa empírica do mundo real

Análises de rejeição de amostras

. $\hat{\mathbf{P}}(X|\mathbf{e}) = \alpha \mathbf{N}_{PS}(X, \mathbf{e})$ (algoritmo defn)

$= \mathbf{N}_{PS}(X, \mathbf{e}) / N_{PS}(\mathbf{e})$ (normalizado por $N_{PS}(\mathbf{e})$)

$\approx \mathbf{P}(X, \mathbf{e}) / P(\mathbf{e})$ (propriedade de *PRIORSAMPLE*)

$= \mathbf{P}(X|\mathbf{e})$,(def. De probabilidade condicional)

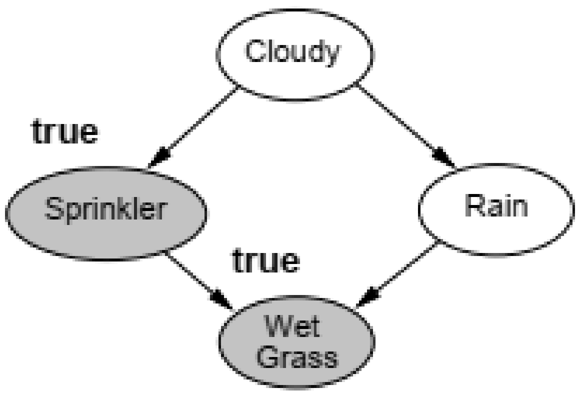
Assim, a rejeição de amostras retorna estimativas posteriores consistentes.

Problema: desesperadamente caro se $P(\mathbf{e})$ for pequeno

Exemplo de ponderação da probabilidade

Estimativa $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$

P(C) = .5



C	P(S)
T	.10
F	.50

C	P(R)
T	.80
F	.20

S	R	P(W)
T	T	.99
T	F	.90
F	T	.90
F	F	.00

Exemplo de ponderação da probabilidade cont

Processo de geração de amostras:

1. $w \leftarrow 1.0$
2. Amostra $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$; diz *true*
3. $Sprinkler$ *true* , então
$w \leftarrow w \times P(Sprinkler = true | Cloudy = true) = 0.1$
4. Amostra $\mathbf{P}(Rain | Cloudy = true) = \langle 0.8, 0.2 \rangle$; diz *true*
5. $WetGrass$ tem valor *true* , então
$w \leftarrow w \times P(WetGrass = true | Sprinkler = true, Rain = true) = 0.099$

Analise de Ponderação de Probabilidade

A probabilidade de amostragem para *WEIGHTEDSAMPLE* é

$$S_{WS}(\mathbf{y},\mathbf{e}) = \Pi_{i=1}^{l} P(y_i|Parents(Y_i))$$

Nota: presta atenção à evidência apenas em $ancestors$

$\Longrightarrow$ em algum lugar "entre" a distribuição prévia e posterior

Peso para uma determinada amostra y, e é

$$w(\mathbf{y},\mathbf{e}) = \Pi_{i=1}^{m} P(e_i|Parents(E_i))$$

A probabilidade de amostragem ponderada é

$$\begin{aligned} & S_{WS}(\mathbf{y},\mathbf{e})w(\mathbf{y},\mathbf{e}) \\ & \quad = \Pi_{i=1}^{l} P(y_i|Parents(Y_i)) \;\; \Pi_{i=1}^{m} P(e_i|Parents(E_i)) \\ & \quad = P(\mathbf{y},\mathbf{e}) \end{aligned}$$

(por semântica global padrão de rede)

Portanto, a ponderação de probabilidade de retorno estimativas consistentes, mas o desempenho ainda se degrada com muitas variáveis de evidência

Inferência aproximada usando MCMC

"State" da rede = atribuição atual a todas as variáveis

Gerar o próximo estado por amostragem de uma variável dada a manta de Markov

Amostra cada variável por sua vez, mantendo as evidências fixas

```
function MCMC-Ask(X,e,bn,N) returns an approximation to P(X|e)
    local variables: N[X], a vector of counts over X, initially zero
                     Y, the nonevidence variables in bn
                     x, the current state of the network, initially copied from e

    initialize x with random values for the variables in Y
    for j = 1 to N do
        N[x] ← N[x] + 1 where x is the value of X in x
        for each Y_i in Y do
            sample the value of Y_i in x from P(Y_i|MB(Y_i)) given the values of MB(Y_i) in x
    return Normalize(N[X])
```

Aproxima-se de uma distribuição estacionária: a fração de tempo de longo prazo gasta em cada estado é exatamente proporcional à sua probabilidade posterior

Exemplo MCMC

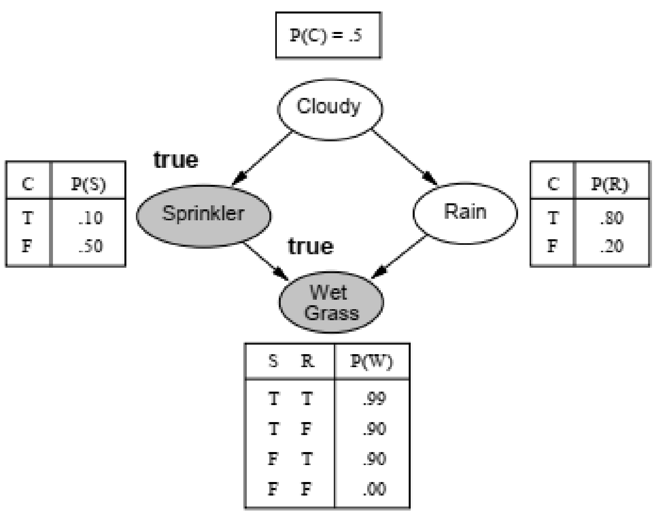
Estimativa $\mathbf{P}(Rain|Sprinkler=true, WetGrass=true)$

Amostra $Cloudy$ então $Rain$, repete .

Número de contagem de tempo $Rain$ é verdadeiro e falso na amostra

Manta de Markov de $Cloudy$ é $Sprinkler$ e $Rain$

Manta de Markov de $Rain$ é $Cloudy$, $Sprinkler$ e $WetGrass$



Exemplo MCMC cont.

Estado inicial aleatório: $Cloudy = true$ e $Rain = false$

1. $\mathbf{P}(Cloudy|MB(Cloudy)) = \mathbf{P}(Cloudy|Sprinkler, \neg Rain)$
   amostra → false
2. $\mathbf{P}(Rain|MB(Rain)) = \mathbf{P}(Rain|\neg Cloudy, Sprinkler, WetGrass)$
   amostra → true

Visite 100 estados

31 tem $Rain = true$, 69 tem $Rain = false$

$$\hat{\mathbf{P}}(Rain|Sprinkler = true, WetGrass = true)$$
$$= \text{NORMALIZE}(\langle 31, 69 \rangle) = \langle 0.31, 0.69 \rangle$$

Análise MCMC: Resumo

Probabilidade de transição $q(\mathbf{y} \to \mathbf{y}')$

Probabilidade de ocupação $\pi_t(\mathbf{y})$ no momento t

A condição de equilíbrio em π_t define a distribuição estacionária $\pi(\mathbf{y})$

Nota: a distribuição estacionária depende da escolha de $q(\mathbf{y} \to \mathbf{y}')$

O equilíbrio detalhado dos estados em pares garante o equilíbrio

Probabilidade de transição de amostragem de Gibbs:
amostragem de cada variável, dados os valores atuais de todas as outras
$\Longrightarrow$balanço detalhado com o verdadeiro posterior

Para redes Bayesianas, a amostragem Gibbs reduz a amostragem condicionada na manta Markov de cada variável

Distribuição estacionária

$\pi_t(\mathbf{y}) =$ probabilidade no estado *y* no tempo *t*

$\pi_{t+1}(\mathbf{y}') =$ probabilidade no estado $\mathbf{y}'$ no tempo *t+1*

π_{t+1} em termos de π_t e $q(\mathbf{y} \to \mathbf{y}')$

$$\pi_{t+1}(\mathbf{y}') = \Sigma_{\mathbf{y}} \pi_t(\mathbf{y}) q(\mathbf{y} \to \mathbf{y}')$$

Distribuição estacionária: $\pi_t = \pi_{t+1} = \pi$

$\pi(\mathbf{y}') \overset{\pi}{=} \Sigma_{\mathbf{y}} \pi(\mathbf{y}) q(\mathbf{y} \to \mathbf{y}')$ para todo $\mathbf{y}'$

Se π existe, e é único (específico para $q(\mathbf{y} \to \mathbf{y}')$)

Em equilíbrio, esperado “outflow” = esperado “inflow”.

Balanço Detalhado

"Outflow" = "inflow" para cada par de estados:

$\pi(\mathbf{y})q(\mathbf{y} \to \mathbf{y}') = \pi(\mathbf{y}')q(\mathbf{y}' \to \mathbf{y})$ para todos $\mathbf{y}$, $\mathbf{y}'$

Balanço detalhado $\Longrightarrow$ estacionário:

$$\begin{aligned} \Sigma_{\mathbf{y}}\pi(\mathbf{y})q(\mathbf{y} \to \mathbf{y}') &= \Sigma_{\mathbf{y}}\pi(\mathbf{y}')q(\mathbf{y}' \to \mathbf{y}) \\ &= \pi(\mathbf{y}')\Sigma_{\mathbf{y}}q(\mathbf{y}' \to \mathbf{y}) \\ &= \pi(\mathbf{y}') \end{aligned}$$

Os algoritmos MCMC normalmente são construídos projetando uma probabilidade de transição q que está em equilíbrio detalhado com o desejado π

Amostras Gibbs

Amostra cada variável por sua vez, dadas *todas as outras variáveis*

Amostra Y_i, deixe $\bar{\mathbf{Y}}_i$ ser todas as outras variáveis não evidenciais
Os valores atuais são y_i e $\bar{\mathbf{y}}_i$; **e** é fixa
A probabilidade de transição é dada por

$$q(\mathbf{y} \to \mathbf{y}') = q(y_i, \bar{\mathbf{y}}_i \to y_i', \bar{\mathbf{y}}_i) = P(y_i' | \bar{\mathbf{y}}_i, \mathbf{e})$$

Isto dá um equilíbrio detalhado com o verdadeiro posterior $P(\mathbf{y}|\mathbf{e})$:

$$\begin{aligned}\pi(\mathbf{y})q(\mathbf{y} \to \mathbf{y}') &= P(\mathbf{y}|\mathbf{e})P(y_i'|\bar{\mathbf{y}}_i, \mathbf{e}) = P(y_i, \bar{\mathbf{y}}_i|\mathbf{e})P(y_i'|\bar{\mathbf{y}}_i, \mathbf{e}) \\ &= P(y_i|\bar{\mathbf{y}}_i, \mathbf{e})P(\bar{\mathbf{y}}_i|\mathbf{e})P(y_i'|\bar{\mathbf{y}}_i, \mathbf{e}) \quad \text{(regra da cadeia)} \\ &= P(y_i|\bar{\mathbf{y}}_i, \mathbf{e})P(y_i', \bar{\mathbf{y}}_i|\mathbf{e}) \quad \text{(regras da corrente ao contrário)} \\ &= q(\mathbf{y}' \to \mathbf{y})\pi(\mathbf{y}') = \pi(\mathbf{y}')q(\mathbf{y}' \to \mathbf{y})\end{aligned}$$

Amostra manto de Markov

Uma variável é independente de todas as outras, dado seu manto Markov:

$$P(y_i'|\bar{\mathbf{y}}_i, \mathbf{e}) = P(y_i'|MB(Y_i))$$

A probabilidade dada pela manto Markov é calculada da seguinte forma:

$$P(y_i'|MB(Y_i)) = P(y_i'|Parents(Y_i))\Pi_{Z_j \in Children(Y_i)} P(z_j|Parents(Z_j))$$

Assim, o cálculo da distribuição de amostras sobre Y_i para cada virada requer apenas multiplicações de cd se Y_i , *c* filhos e *d* valores ; pode armazená-la em cache se *c* não for muito grandes.

Principais problemas computacionais:

1) Difícil dizer se a convergência foi alcançada.

2) pode ser um desperdício se o manto de Maskov for grande:

$P(Y_i|MB(Y_i))$ não terá muitas chances (lei de grandes números)

Performance de algoritmos de aproximação

Aproximação absoluta: $|P(X|\mathbf{e}) - \hat{P}(X|\mathbf{e})| < \epsilon$

Aproximação relativa: $\frac{|P(X|\mathbf{e}) - \hat{P}(X|\mathbf{e})|}{P(X|\mathbf{e})} \leq \epsilon$

Relativa $\Longrightarrow$ absoluta desde $0 \leq P \leq 1$ (pode ser $O(2^{-n})$)

Algoritmos aleatórios podem falhar com probabilidades no máximo δ

Aproximação poli-time: $\text{poly}(n, \epsilon^{-1}, \log \delta^{-1})$

Teorema (Dagum e Luby, 1993): tanto a aproximação absoluta quanto a relativa para algoritmos deterministicos ou randomizados são NP_hard para qualquer $\epsilon, \delta < 0.5$
(Aproximação absoluta de poli-times sem evidência -- limites Chernoff)

Lógica Fuzzy

- É a lógica baseada em análises de informações estritamente qualitativas. Isto é feito de forma que a decisão não se resuma entre um ‘sim’ e um ‘não’, mas, também considera abstrações do tipo ‘próximo de’, ‘em torno de’, ‘muito alto’, ‘bem baixo’, etc.

Lógica Fuzzy

- Exemplo: Homens de meia idade
- Lógica Clássica:
- Se 40 ≤ Idade ≤ 55 então Homem meia idade
- Lógica fuzzy:

Idade	35	40	45	50	55
Grau de pertinência	0,0	0,5	1,0	0,5	0,0

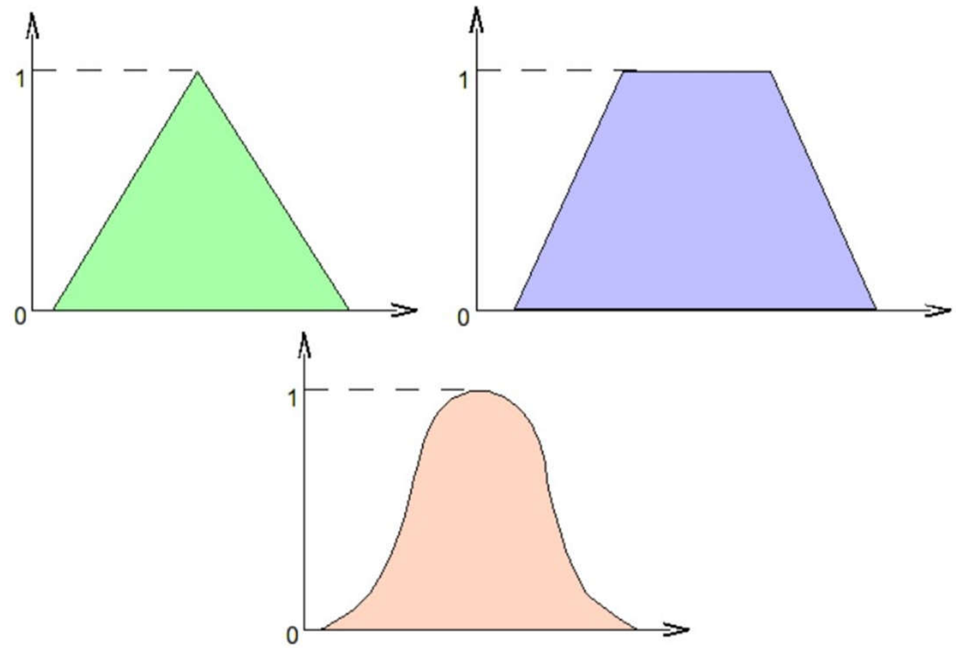
Conjuntos Fuzzy

- Um conjunto fuzzy X em um universo de discurso U é caracterizado por uma função que assume valores no intervalo [0,1]

$$\mu_X(u) \in [0,1], \forall u \in U$$

Conjuntos Fuzzy

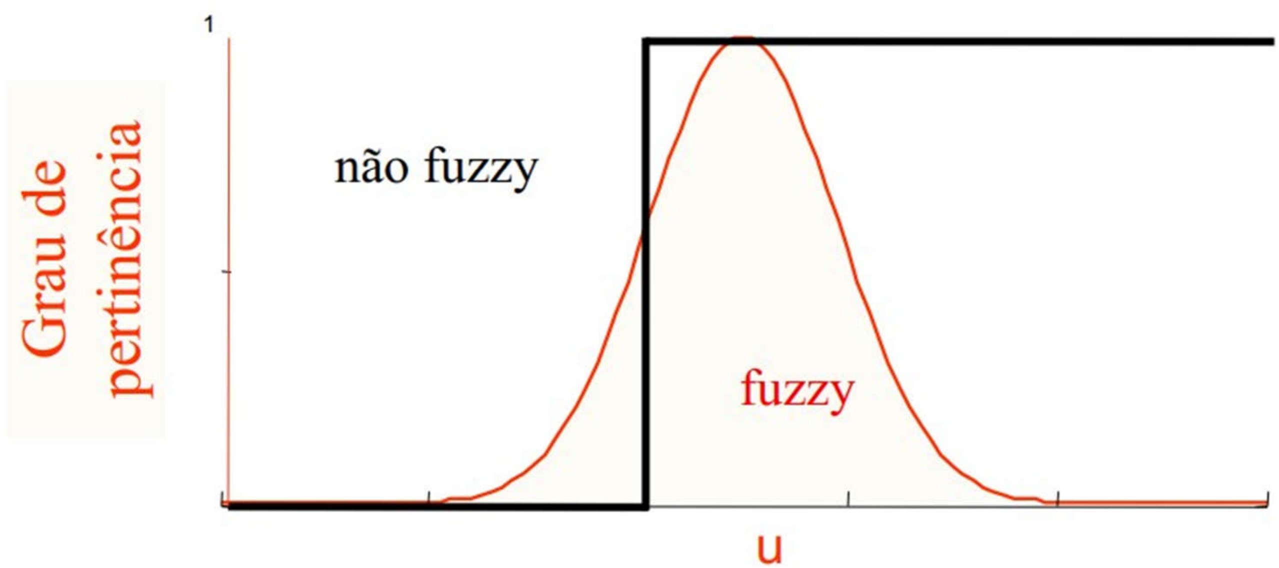
- Funções de pertinência
  - Triangular
  - Trapezoidal
  - Sino
  - Gaussiana
  - Sigmoidal



Conjuntos Fuzzy

- Contínua

$$\mu_F(u) = \exp\left[-\frac{(u-u_0)^2}{2\sigma^2}\right]$$



Conjuntos Fuzzy

- Discreta

$$F = \{(u, \mu_F(u)), u \in U\}$$

- Notação

$$F = \left\{ \frac{\mu_F(u)}{u}, u \in U \right\}$$

$$F = \{\mu_F(u), u \in U\}$$

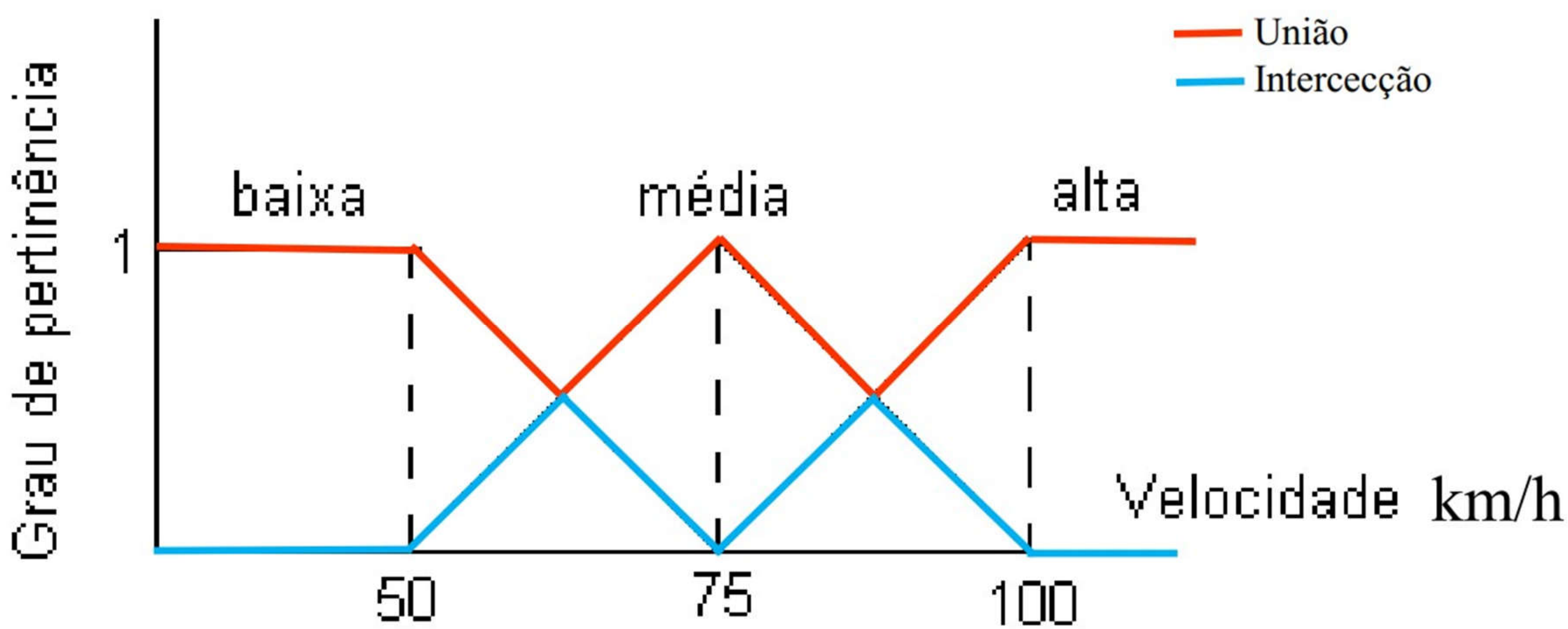
- Exemplo

$$F = \{(30,0),(35,0,3),(40,1),(45,1),(50,0,7),(55,0,4),(60,0)\}$$

$$F = \{(\frac{0}{30}),(\frac{0,3}{35}),(\frac{1}{40}),(\frac{1}{45}),(\frac{0,7}{50}),(\frac{0,4}{55}),(\frac{0}{60})\}$$

$$F = \{0,0.3,1,1,0.7,0.4,0)\}$$

Operações



União
Intercecção
baixa
média
alta
1
Grau de pertinência
Velocidade km/h
50
75
100

Variáveis linguísticas

- Uma variável linguística é uma variável cujos valores são palavras
- Uma variável linguística é definida por

<X, T(X), U, G, M>

- X: nome, T(X):função de pertinência de X, U:universo de discurso, G: gramática, M: regras semânticas associadas

Regras Fuzzy

- Relacionam variáveis fuzzy, cada uma delas associada a um dos seus predicados linguísticos

- SE Velocidade é Baixa
  ENTÃO Aceleração é Alta

Base de conhecimento

- Base de dados: definições de conjuntos fuzzy
- Base de regras
- Exemplo de uma regra SE-ENTÃO:
    - Se Erro é Pequeno e Variação do erro é Baixa então:
    - posição da válvula tampão é ZERO.
    - Parte SE: antecedente
    - Parte ENTÃO: consequente

Inferência e defuzificação

- Operação de max-min
  - Inferência: Operador mínimo
  - Agregação: Operador máximo
- Defuzificação
  - Centro de área