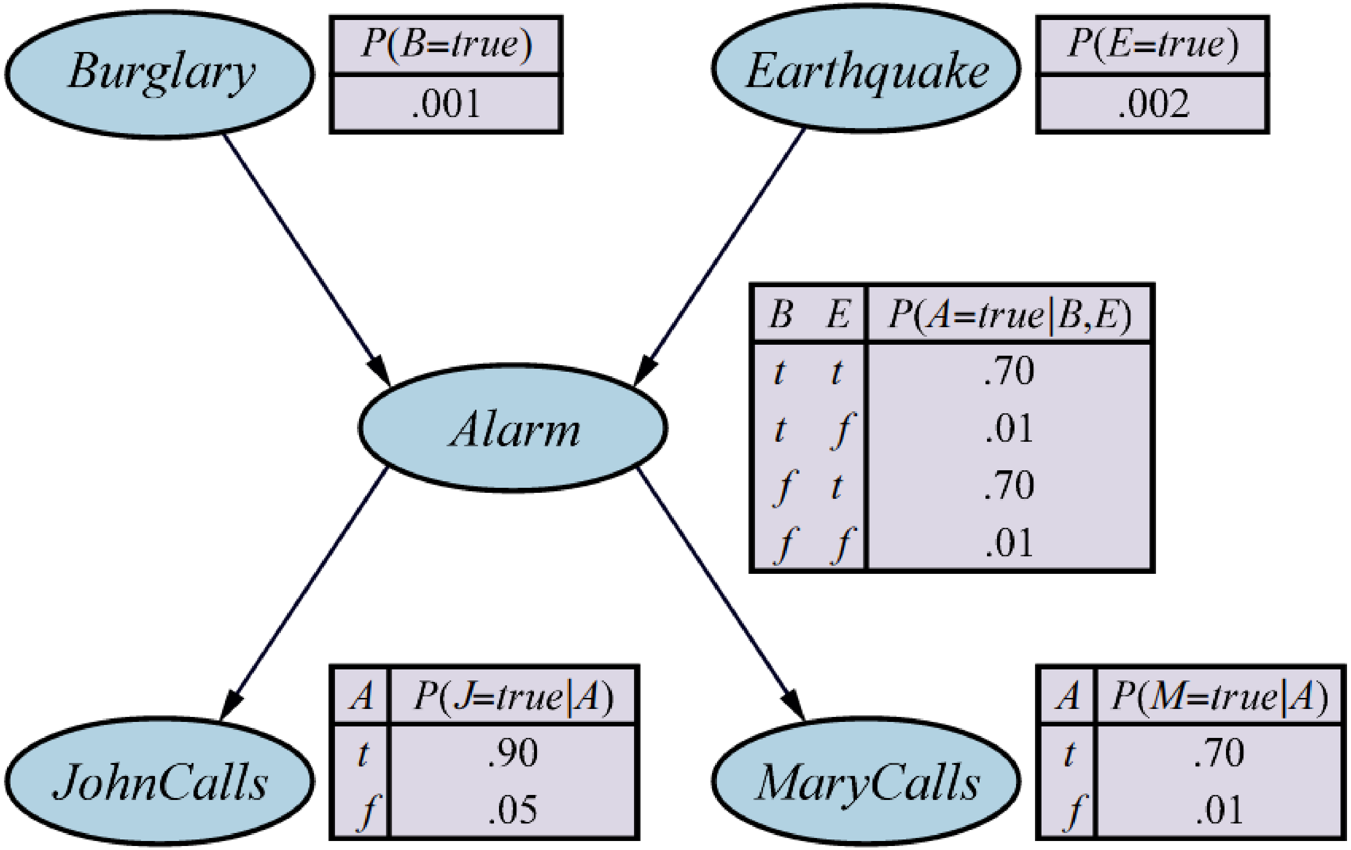
conhecimento incerto - inferência

rede Bayesiana (recap.)



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true|B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

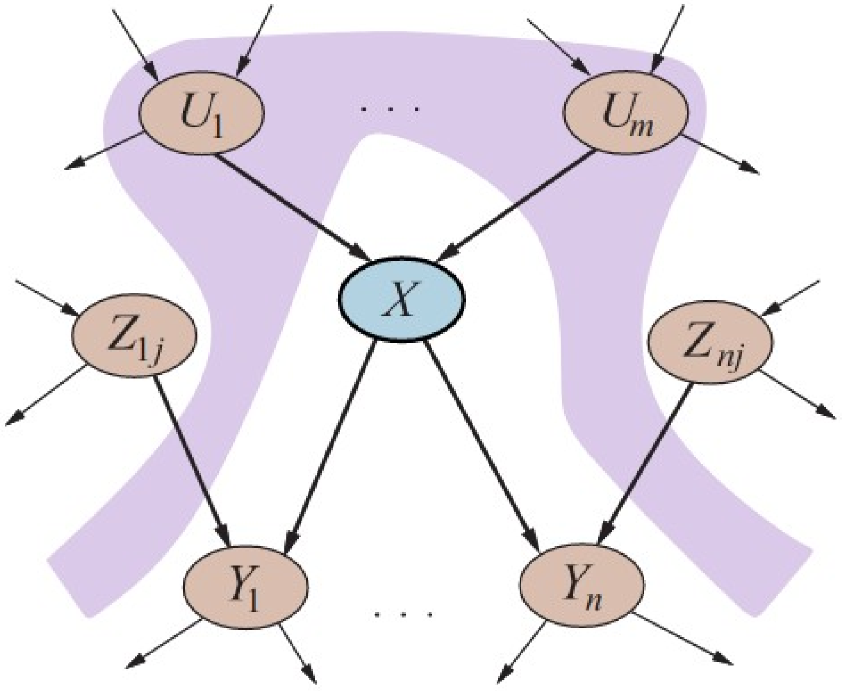
representação eficiente de conhecimento probabilístico

as tabelas de probabilidades condicionadas permitem obter as probabilidades conjuntas

propriedades da rede Bayesiana

uma variável é condicionalmente independente dos seus **não-descendentes** dados os seus pais

ex: *X* é condicionalmente independente dos nós Z_{ij} dados U_i

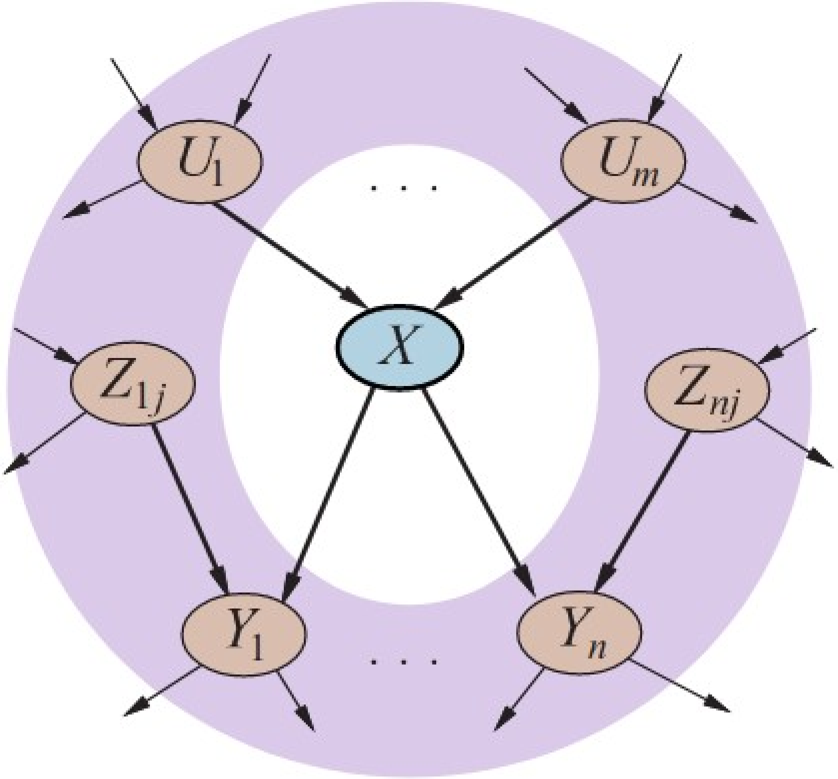


J é independente de *B*, *E* e *M* dado *A*(*larm*)

propriedades II

um nó é condicionalmente independente de todos os outros nós da rede dados os seus pais, descendentes e pais dos descendentes – **manto de Markov**

X é independente de todo o resto da rede dados U_i, Z_{ij} e Y_i



B é independente de *J* e *M* dado *A*(*larm*) e *E*

inferência exata em redes Bayesianas

consiste em:

calcular a probabilidade de variáveis *questionadas* (*query*), dadas variáveis *evidência*

X – variáveis questionadas

E – variáveis evidência

Y – variáveis escondidas (não questionadas, não evidência)

ex: $\mathbf{P}(X \mid \mathbf{e})$

$\mathbf{P}(B \mid J{=}t, M{=}t) = \langle 0{,}284,\ 0{,}716 \rangle$

inferência por enumeração

já vimos que $\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$

e que não escala bem!

ex. da pg. anterior: $\mathbf{P}(B|j, m) = \alpha \mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, j, m, e, a)$

em que *e* e *a*, são os valores das v.a. *E* e *A*

em redes Bayesianas (para *B* = *t*)

$$P(b|j, m) = \alpha \sum_{e} \sum_{a} P(b) P(e) P(a|b, e) P(j|a) P(m|a)$$

complexidade, com *n* v.a. booleanas, $O(n2^n)$

melhorando...

$$P(b|j,m)=\alpha\sum_{e}\sum_{a}P(b)P(e)P(a|b,e)P(j|a)P(m|a)$$

pode rescrever-se

$$P(b|j,m)=\alpha P(b)\sum_{e}P(e)\sum_{a}P(a|b,e)P(j|a)P(m|a)$$

ainda assim tem complexidade temporal $O(2^n)$ para n variáveis booleanas

eliminação de variáveis

algoritmo mais simples que faz as contas apenas uma vez e guarda para uso futuro (em que aparece a mesma expresão)

$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a|B,e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j|a)}_{\mathbf{f}_4(A)} \underbrace{P(m|a)}_{\mathbf{f}_5(A)}$$

de notar que $\mathbf{f}_4(A)$ e $\mathbf{f}_5(A)$ só dependem de A, porque J e M são fixados pela questão

$$\mathbf{f}_4 = \begin{pmatrix} P(j|a) \\ P(j|\neg a) \end{pmatrix} = \begin{pmatrix} 0{,}90 \\ 0{,}05 \end{pmatrix}$$

$$\mathbf{f}_5 = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix} = \begin{pmatrix} 0{,}70 \\ 0{,}01 \end{pmatrix}$$

$\mathbf{f}_3(A, B, E)$ é uma matriz de 2 x 2 x 2
1º elemento $P(a \mid b, e) = 0{,}95$
últ. elem $P(\neg a \mid \neg b, \neg e) = 0{,}999$

Ciências ULisboa

avaliação

produto pontual

usando a fatorização

$$\mathbf{P}(B|j,m)=\alpha\mathbf{f}_1(B)\times\sum_e \mathbf{f}_2(E)\times\sum_a \mathbf{f}_3(A,B,E)\times\mathbf{f}_4(A)\times\mathbf{f}_5(A)$$

da direita para a esquerda somando fora variáveis

$$\mathbf{f}_6(B,E)=\sum_a \mathbf{f}_3(A,B,E)\times\mathbf{f}_4(A)\times\mathbf{f}_5(A)$$

$$=(\mathbf{f}_3(a,B,E)\times\mathbf{f}_4(a)\times\mathbf{f}_5(a))+(\mathbf{f}_3(\neg a,B,E)\times\mathbf{f}_4(\neg a)\times\mathbf{f}_5(\neg a))$$

donde $\mathbf{f}_7(B)=\sum_e \mathbf{f}_2(E)\times\mathbf{f}_6(B,E)=\mathbf{f}_2(e)\times\mathbf{f}_6(B,e)+\mathbf{f}_2(\neg e)\times\mathbf{f}_6(B,\neg e)$

finalmente $\mathbf{P}(B|j,m)=\alpha\mathbf{f}_1(B)\times\mathbf{f}_7(B)$ que pode ser avaliada diretamente

produto pontual

$\times$ é o produto pontual

o produto pontual de dois fatores $\mathbf{f}_1$ e $\mathbf{f}_2$ é um fator $\mathbf{f}$ cujas variáveis são a união das de $\mathbf{f}_1$ e $\mathbf{f}_2$ e os valores são os produtos dos elementos correspondentes em ambos

ex: se as variáveis $Y_1,\ldots,Y_k$ são comuns a $\mathbf{f}_1$ e $\mathbf{f}_2$

$$\mathbf{f}(X_1\ldots X_j, Y_1\ldots Y_k, Z_1\ldots Z_l) = \mathbf{f}_1(X_1\ldots X_j, Y_1\ldots Y_k)\,\mathbf{f}_2(Y_1\ldots Y_k, Z_1\ldots Z_l)$$

produto pontual (ex.)

X	Y	$\mathbf{f}(X,Y)$	Y	Z	$\mathbf{g}(Y,Z)$	X	Y	Z	$\mathbf{h}(X,Y,Z)$
t	t	$.3$	t	t	$.2$	t	t	t	$.3 \times .2 = .06$
t	f	$.7$	t	f	$.8$	t	t	f	$.3 \times .8 = .24$
f	t	$.9$	f	t	$.6$	t	f	t	$.7 \times .6 = .42$
f	f	$.1$	f	f	$.4$	t	f	f	$.7 \times .4 = .28$
						f	t	t	$.9 \times .2 = .18$
						f	t	f	$.9 \times .8 = .72$
						f	f	t	$.1 \times .6 = .06$
						f	f	f	$.1 \times .4 = .04$

multiplicam-se os elementos com o valor de Y em comum

somar fora uma variável

somar as sub-matrizes com cada valor da variável

ex:

$$\mathbf{f}(B,C)=\sum_a \mathbf{f}_3(A,B,C)=\mathbf{f}_3(a,B,C)+\mathbf{f}_3(\neg a,B,C)$$

$$=\begin{pmatrix}0{,}06 & 0{,}24\\ 0{,}42 & 0{,}28\end{pmatrix}+\begin{pmatrix}0{,}18 & 0{,}72\\ 0{,}0{,}6 & 0{,}04\end{pmatrix}=\begin{pmatrix}0{,}24 & 0{,}96\\ 0{,}48 & 0{,}32\end{pmatrix}$$

reordenam-se as variáveis de modo a ficar no somatório só com as que vão ser somadas
ex, somar fora *E* em

$$\sum_e \mathbf{f}_2(E)\times\mathbf{f}_3(A,B,E)\times\mathbf{f}_4(A)\times\mathbf{f}_5(A)=\mathbf{f}_4(A)\times\mathbf{f}_5(A)\times\sum_e \mathbf{f}_2(E)\times\mathbf{f}_3(A,B,E)$$

variáveis irrelevantes para a questão

supondo a questão $\mathbf{P}(JohnCalls|Burglary=true)$

$$\mathbf{P}(J|b)=\alpha P(b)\sum_{e} P(e)\sum_{a} P(a|b,e)\mathbf{P}(J|a)\sum_{m} P(m|a)$$

verifica-se de imediato $\sum_{m} P(m|a)=1$

e pode aplicar-se:

qualquer variável que não seja antecedente da variável questionada, ou uma variável de evidência, é irrelevante para a questão e pode ser eliminada antes de avaliar a questão

complexidade da inferência exata

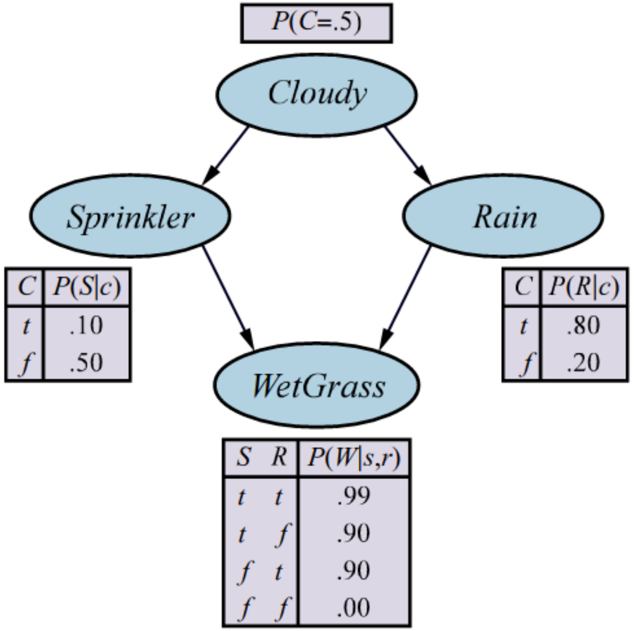
polytrees – há no máximo um caminho entre dois nós

nestas estruturas a eliminação de variáveis tem complexidade espacial e temporal lineares com a dimensão da rede

redes multiplamente conectadas

eliminação de variáveis pode ter complexidade temporal e espacial exponenciais

mesmo com número de pais por nó limitado!



Ciências ULisboa

inferência aproximada em redes Bayesianas

visto o problema da inferência exata ser intratável para redes grandes…

fazem-se amostragens de Monte Carlo (aleatórias)

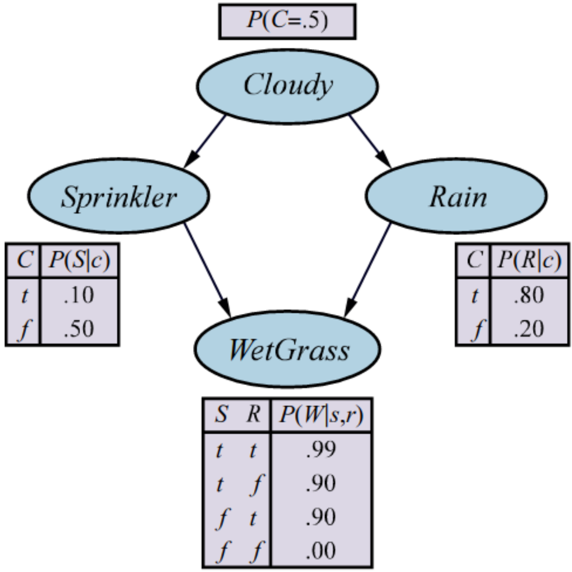
dois métodos práticos:

- amostragem direta
- amostragem por cadeia de Markov

amostragem direta

amostrar cada variável em ordem topológica

a distribuição de cada amostragem é condicionada aos valores já atribuídos aos pais



ex:

1. amostra de **P**(*Cloudy*)=⟨0,5; 0,5⟩, valor *true*
2. amostra de **P**(*Sprinkler*|*Cloudy=true)*=⟨0,1; 0,9⟩, valor *false*
3. amostra de **P**(*Rain*|*Cloudy=true)*=⟨0,8; 0,2⟩, valor *true*
4. amostra de **P**(*WetGrass*|*Sprinkler=false, Rain=true)*=⟨0,9; 0,1⟩, valor *true*

amostra retorna [*true*, *false*, *true*, *true*]

probabilidade estimada

probabilidade de amostrar evento $(x_1,\ldots,x_n)$:

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

contam-se os eventos N e:

$$\lim_{N\to\infty}\frac{N_{PS}(x_1,\ldots,x_n)}{N}=S_{PS}(x_1,\ldots,x_n)=P(x_1,\ldots,x_n)$$

resultado da amostragem $\dfrac{N_{PS}(x_1,\ldots,x_n)}{N}=\hat{P}(x_1,\ldots,x_n)\approx P(x_1,\ldots,x_n)$

Ciências ULisboa

amostragem com rejeição

pretende-se determinar $\mathbf{P}(X|\mathbf{e})$

obtêm-se amostras da rede

rejeitam-se as que não condizem com a evidência

estimativa: $\hat{\mathbf{P}}(X|\mathbf{e}) = \frac{\mathbf{N}_{PS}(X,\mathbf{e})}{N_{PS}(\mathbf{e})}$

rejeição cresce exponencialmente com v.a.s de evidência!

ex: obter $\hat{\mathbf{P}}(Rain|Sprinkler=true)$ com 100 amostras

em 73 delas *Sprinkler=false*, são rejeitadas, das 27 restantes

Rain=true em 8 e *Rain=false* em 19, logo

$$\hat{\mathbf{P}}(R|S=t) = \alpha(\langle 8;19\rangle) = \langle 0{,}296;0{,}704\rangle$$

Ciências ULisboa

Markov chain Monte Carlo (MCMC)

em vez de gerar cada amostra desde o início,

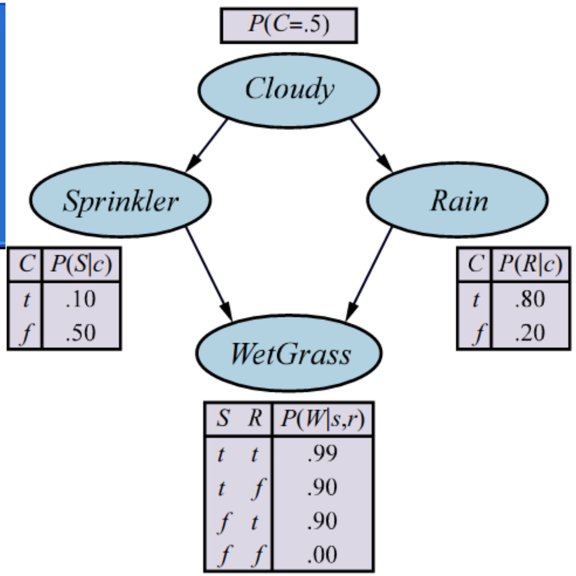
cada amostra é uma pequena alteração da anterior

- a amostragem Gibbs (uma forma de MCMC)

  começa num estado arbitrário com as v.a. de evidência fixadas nos valores observados

  gera um novo estado amostrando aleatoriamente uma das variáveis não evidência, X_i, condicionado ao valor corrente das v.a. do manto de Markov de X_i

ex. Gibbs MCMC



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

C S R W

são fixadas: *S*=*t* e *W*=*t* estado inicial: [*t*, *t*, *f*, *t*]

e incializadas aleatoriamente: *C*=*t* e *R*=*f*

e C e R amostradas repetidamente em ordem aleatória

1. *C* amostrada dados os valores do manto de Markov
   i.e. de **P**(*C* | *S*=*t*, *R*=*f*). Supondo *C*=*f*, o estado é [*f*, *t*, *f*, *t*]
2. *R* amostrada dados os valores do manto de Markov
   i.e. de **P**(*R* | *C*=*f*, *S*=*t*, *W*=*t*). Supondo *R*=*t*, o estado é [*f*, *t*, *t*, *t*]

cada estado é uma amostra
se em 20 *R*=*t* e em 60 *R*=*f*, então $\hat{\mathbf{P}}(R | s, w) = \alpha(\langle 20; 60 \rangle) = \langle 0{,}25; 0{,}75 \rangle$

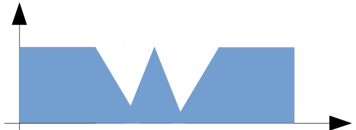
outros modelos de incerteza

teoria de Dempster-Shafer

função de crença (*Belief*): $Bel(A) + Bel(\neg A) \leq 1$

uma versão pessimista de probalilidade…

conjuntos difusos e lógica difusa



representação de pertença em categorias de limites definidos de uma forma contínua em vez de em degraus disjuntos

e inferência sobre essas representações