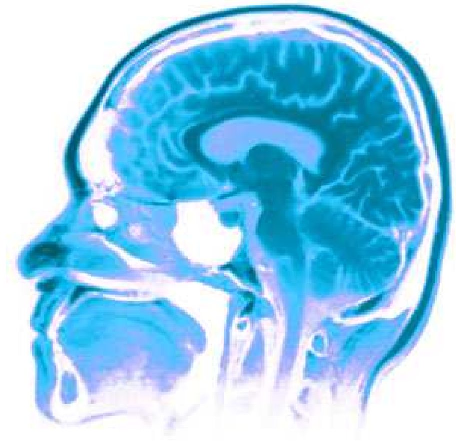
CPSC340



Inference in Probabilistic Graphical Models

Nando de Freitas
September, 2012
University of British Columbia

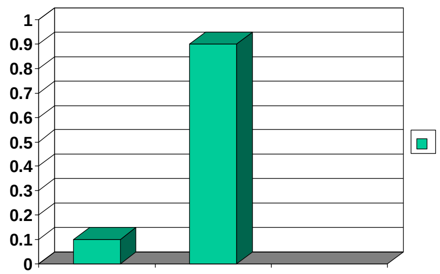
Outline of the lecture

This lecture is devoted to the problem of inference in probabilistic graphical models (aka Bayesian nets). The goal is for you to:

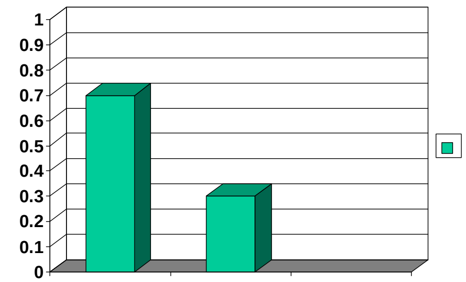
- Practice marginalization and conditioning.
- Learn how to apply dynamic programming to do inference.

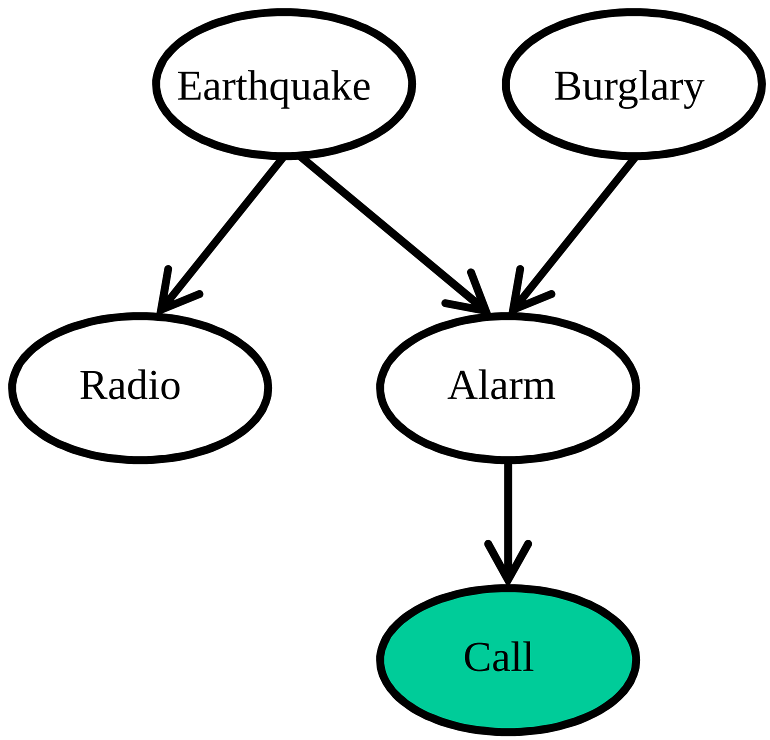
Inference

$P(E=t|C=t)=0.1$



$P(B=t|C=t) = 0.7$

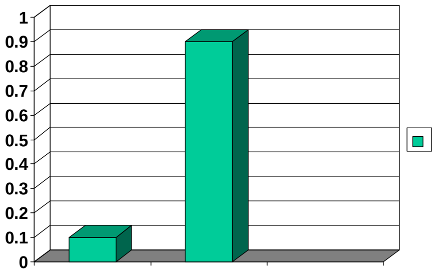




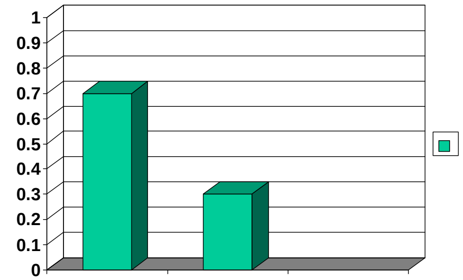
C=t

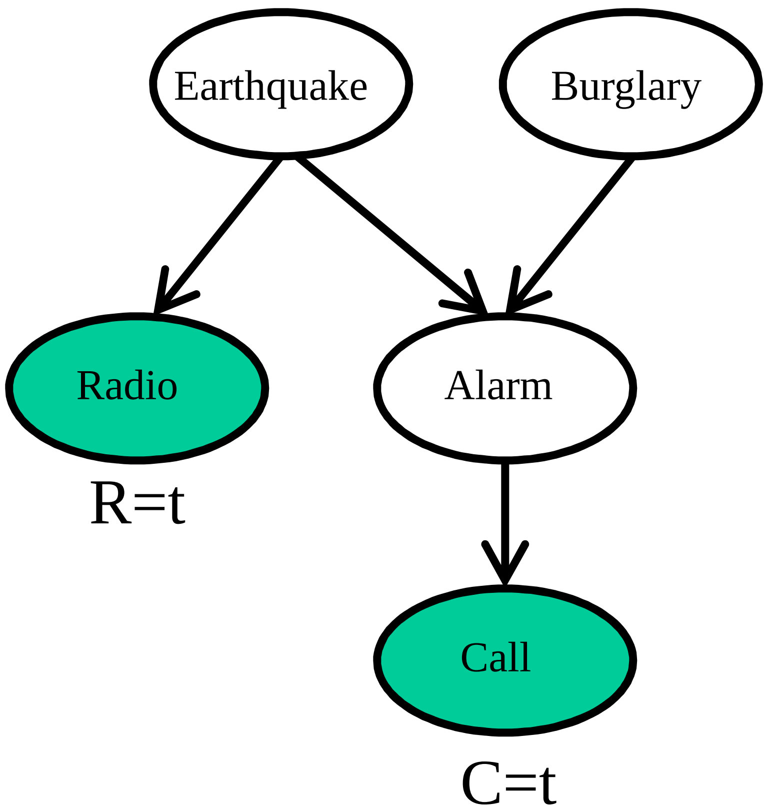
Inference

P(E=t|C=t)=0.1



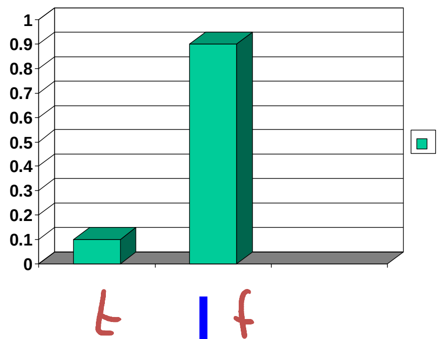
P(B=t|C=t) = 0.7



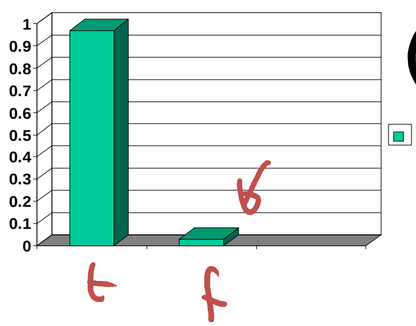


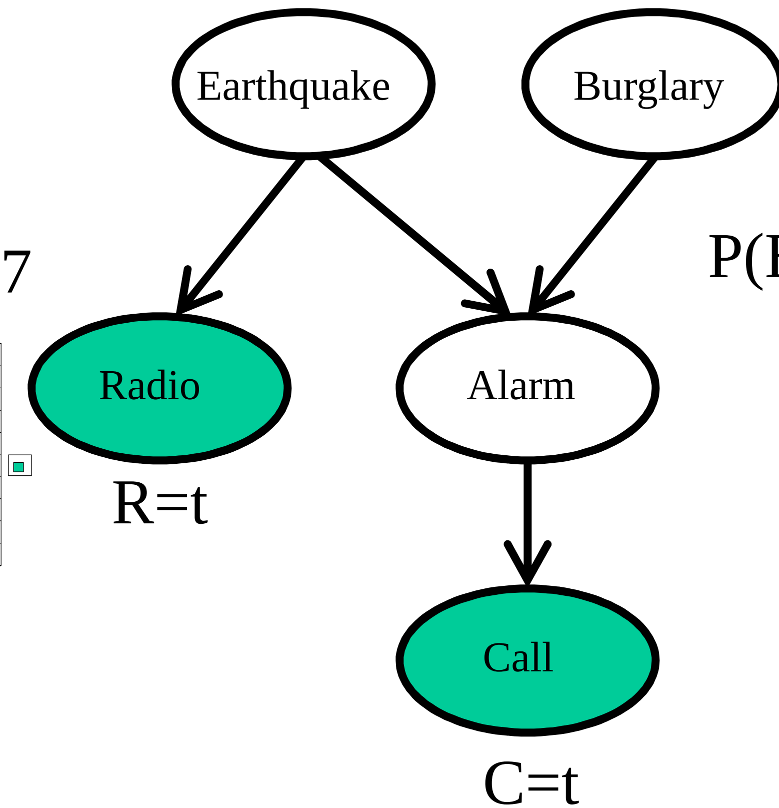
Inference

P(E=t|C=t)=0.1

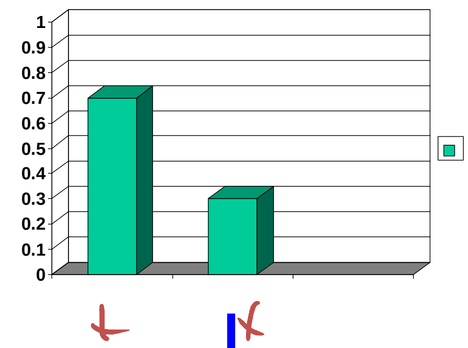


P(E=t|C=t,R=t)=0.97

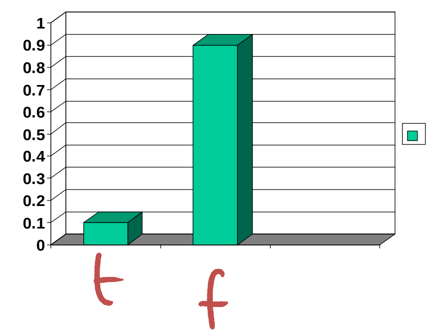




P(B=t|C=t) = 0.7



P(B=t|C=t,R=t) = 0.1



Explaining away effect

The sprinkler network

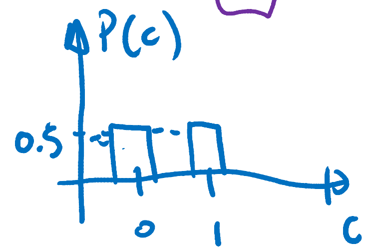
$$P(c) = \begin{array}{c|c} 0 & 1 \\ \hline 0.5 & 0.5 \end{array} \qquad P(c=0) = 0.5$$

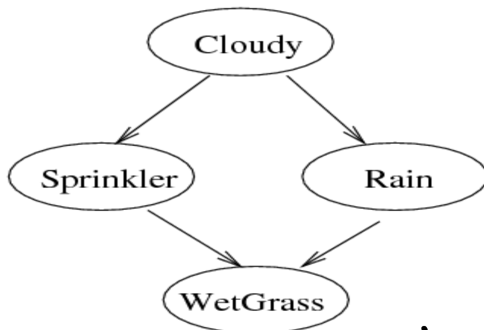
$$P(s|c=1) = \begin{array}{c|c} 0 & 1 \\ \hline 0.9 & 0.1 \end{array}$$

$$\sum_s P(s|c=1) = 1$$

P(c)

P(C=F)	P(C=T)
0.5	0.5



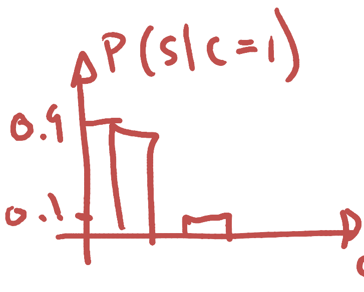


P(s|c)

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

$$P(c=0, s=0) = P(s=0|c=0)P(c=0)$$

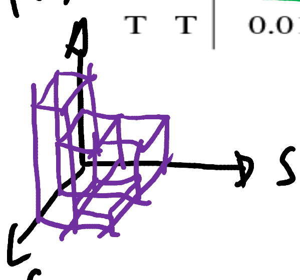


S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

$$P(w|s=0, r=1) = \begin{array}{c|c} 0 & 1 \\ \hline 0.1 & 0.9 \end{array}$$

$$P(c,s) = \begin{array}{c|cc} & s=0 & s=1 \\ \hline c=0 & .25 & .25 \\ c=1 & 0.45 & 0.05 \end{array}$$

p(c,s)



$$p(C, S, R, W) = p(C)p(S|C)p(R|C)p(W|S, R)$$

Inference in DAGs

Let us use 0 to denote false and 1 to denote true.

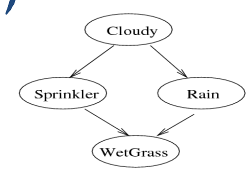
What is the ***marginal probability****,* ***P(S=1)****, that the sprinkler is on?*

$$P(S=1) = \sum_{C=0}^{1}\sum_{R=0}^{1}\sum_{W=0}^{1} P(C,R,W,S=1)$$

$$= \sum_{C}\sum_{R}\sum_{W} P(C)P(S=1|C)P(R|C)P(W|S=1,R)$$

$P(C)$

P(C=F)	P(C=T)
0.5	0.5



$P(S|C)$

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

$P(R|C)$

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

$P(W|R,S)$

Brute force (exponential) approach

*What is the **marginal probability**, **P(S=1)**, that the sprinkler is on?*

$$\sum_{C}\sum_{R}\sum_{W=0}^{1} P(W|S=1,R)\,P(S=1|C)\,P(R|C)\,P(C)$$

$ab + ac$

$a(b+c)$

$$= \overset{000}{} \; P(W=0|S=1,R=0) \times P(S=1|C=0) \times P(R=0|C=0) \times P(C=0)$$

$$ \;\; 0.1 \qquad\qquad 0.5 \qquad\qquad 0.8 \qquad\qquad 0.5$$

$$\overset{001}{+} \; P(W=1|S=1,R=0)\,P(S=1|C=0)\,P(R=0|C=0)\,P(C=0)$$

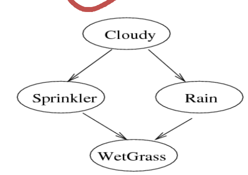
$$\overset{010}{+} \; P(W=0|S=1,R=1)\,P(S=1|C=0)\,P(R=1|C=0)\,P(C=0)$$

$$\overset{011}{+} \; \cdots$$

$$\vdots$$

$$\overset{111}{+}$$

P(C=F)	P(C=T)
0.5	0.5



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S R	P(W=F)	P(W=T)
F F	1.0	0.0
T F (10)	0.1	0.9
F T	0.1	0.9
T T	0.01	0.99

Brute force (exponential) approach

*What is the **marginal probability**, **P(S=1)**, that the sprinkler is on?*

```
PROD
FOR R=0:1:1
    FOR C=0:1:1
        FOR W=0:1:1
            PROD = PROD + P(C)P(R|C)P(S=1|C)P(W|S=1,R)
        END
    END
END
```

Smart approach: variable elimination, *aka* dynamic programming, *aka* distributive law

What is the ***marginal probability***, ***P(S=1)***, *that the sprinkler is on?*

$$P(S=1) = \sum_C \sum_R \sum_W P(W|S=1,R)\, P(R|C)\, P(C)\, P(S=1|C)$$

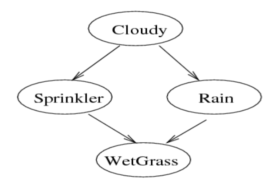
$$= \sum_C \sum_R P(R|C)\, P(C)\, P(S=1|C) \boxed{\cancelto{1}{\sum_W P(W|S=1,R)}} \sim \phi = 1$$

$$= \sum_C P(C)\, P(S=1|C) \boxed{\cancelto{1}{\sum_R P(R|C)}} \sim \psi = 1$$

$$= P(S=1|C=0)\, P(C=0) + P(S=1|C=1)\, P(C=1)$$

$$= 0.3$$

P(C=F)	P(C=T)
0.5	0.5



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

Smart approach: variable elimination, *aka* dynamic programming, *aka* distributive law

What is the ***marginal probability****,* ***P(S=1)****, that the sprinkler is on?*

$\psi = 0$

$\phi = 0$

$\Theta = 0$

FOR $W = 0:1:1$

$\quad \phi_R = \phi_R + P(W \mid S=1, R)$

END

FOR $R = 0:1:1$

$\quad \psi_C = \psi_C + P(R \mid C)\,\phi_R$

END

FOR $C = 0:1:1$

$\quad \Theta = \Theta + P(S=1 \mid R)\,P(C)\,\psi_C \longrightarrow \Theta = 0.3$

END

Smart approach: variable elimination, *aka* dynamic programming, *aka* distributive law

We won't implement the general code in this course. To do this one needs to learn about the ***junction tree*** *data structure. This structure, once created, enables us to conduct any query on the graph very efficiently.*

These exact algorithms work well for small graphs and for graphs that are ***trees*** *or close to trees (have low tree-width). For large densely connected graphs we require the use of algorithms beyond the scope of this course. One of those algorithms is called* ***Gibbs sampling***.

In the next class, we will address a tree of great interest: a chain of nodes. For discrete models this model is known by the name ***HMM***.

Inference in DAGs

What is the ***posterior probability****,* ***P(S=1|W=1),*** *that the sprinkler is on given that the grass is wet?*

$$P(S=1 \mid W=1) = \frac{P(S=1, W=1)}{P(W=1)}$$

$$P(W=1) = \sum_{S} \sum_{C} \sum_{R} P(S, W=1, C, R)$$

$$P(S=1, W=1) = \sum_{C} \sum_{R} P(S=1, W=1, C, R)$$

Inference in DAGs

What is the ***posterior probability***, ***P(S=1|W=1,R=1)***, *that the sprinkler is on given that the grass is wet and it is raining?*

$$P(S=1|W=1,R=1) = P(S=1 \mid (W=1,R=1))$$

$$= \frac{P(S=1,W=1,R=1)}{P(W=1,R=1)}$$

Next lecture

In the next lecture, we will learn to do inference on a tree-structured graphical model, known as hidden Markov model (HMM)