Chapter 5 – Uncertainty and Reasoning

5.2 Probabilistic Reasoning

Bayesian Networks

Outline


- Syntax
- Semantics
- Parameterized distributions

Bayesian Networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions.

Syntax:

- a set of nodes, one per variable
- a directed, acyclic graph (link ≈ "directly influences")
- a conditional distribution for each node given its parents:

$$P(X_i \;\; Parents(X_i))$$

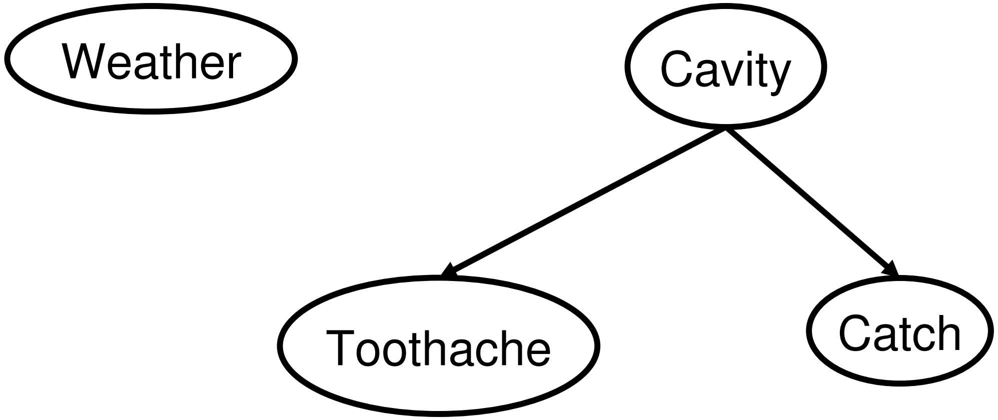
In the simplest case, conditional distribution represented as

a conditional probability table (CPT) giving the

distribution over X_i for each combination of parent values.

Example 1

Topology of network encodes conditional independence assertions:



Weather is independent of the other variables

Toothache and Catch are conditionally independent given Cavity


Example 2

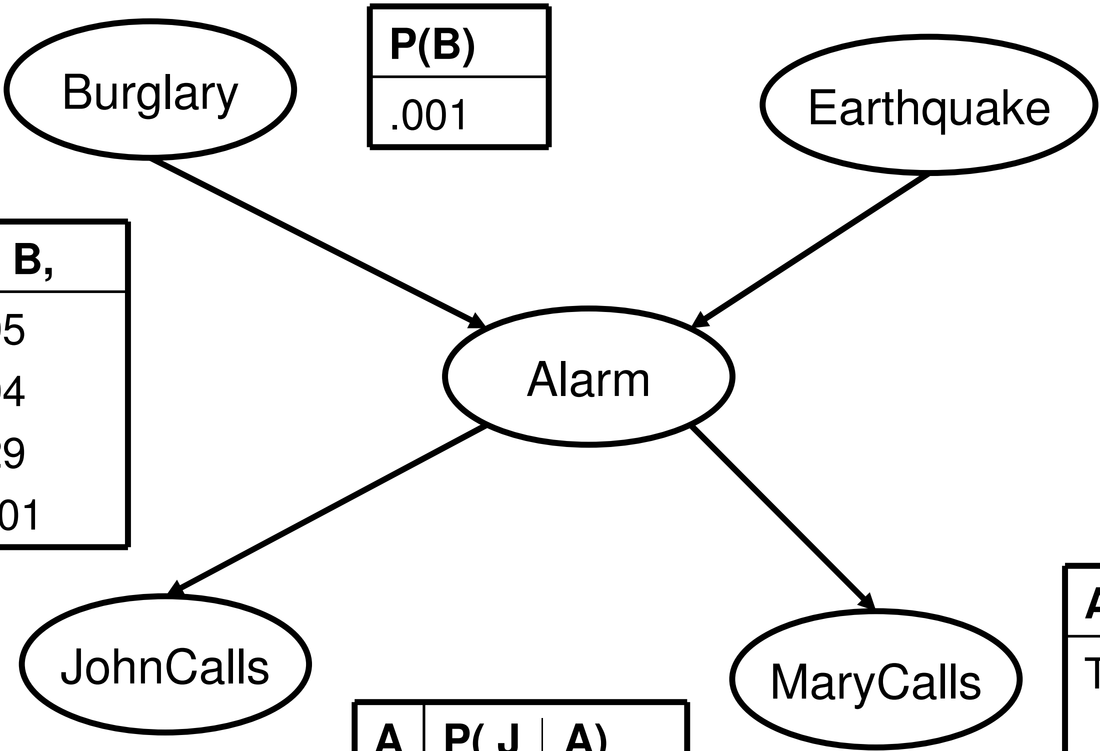
I’m at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn’t call. Sometimes it’s set off by minor earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects “causal” knowledge:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

Example 2 continued.



P(B)
.001

P(E)
.002

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

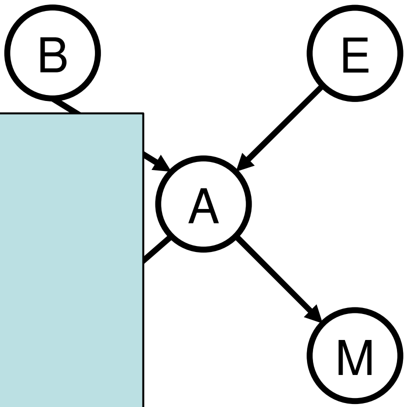
A	P(J | A)
T	.90
F	.05

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

Compactness

A CPT for Boolean X_i with k Boolean parents has

2^k rows for the combinations of parent values



Each row require

(the number for

If each variable

the complete ne

I.e., grows linea

For burglary ne

One more Example:

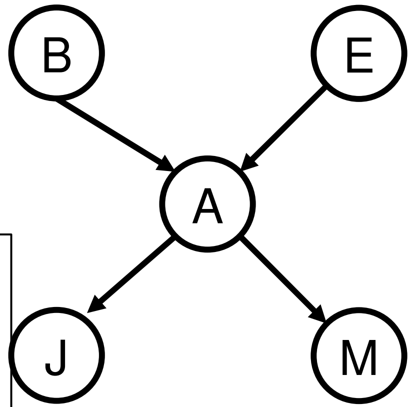
say n nodes n = 30
With k parents each k = 5

Bayesian Network = 960 nodes

Full Joint Distribution > one billion nodes !!

Global Semantics

"Global" semantics defines the full joint distribution as the product of the local conditional distributions:



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

For Example:

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$

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

= 0.90 * 0.70 * 0.001 * 0.999 *o.998 = 0.00062

Local Semantics

Local semantics: each node is conditionally independent of its non descendants given its parents

Example:

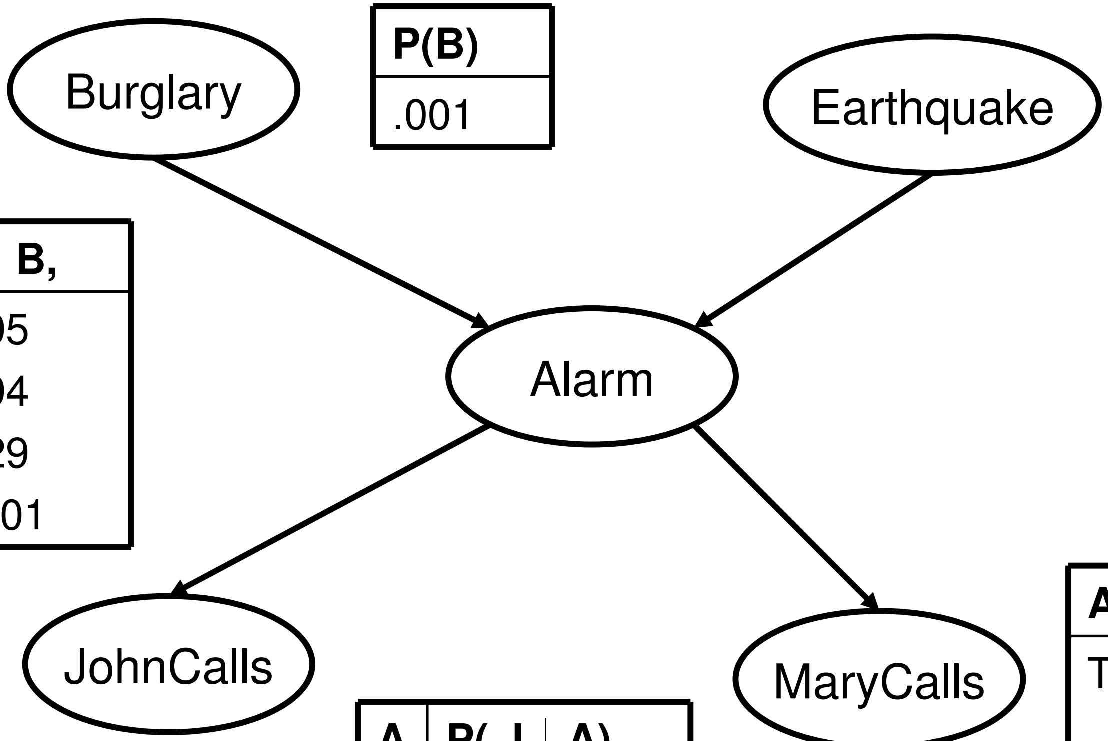
JohnCalls is conditionally independent of

Burglary and *Earthquake*

Given the value of *Alarm*

Theorem: Local semantics <=> global semantics

Example 2 continued.



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

Markov Blanket

Each node is conditionally independent of all others given its Markov blanket: parents + children + children's parents

Example:

Burglary is independent of *JohnCalls* and *MaryCalls*

Given *Alarm* and *Earthquake*

Constructing Bayesian Networks

Need a method such that a series of locally testable assertions of conditional guarantees the required global semantics

1. Choose an ordering of variables
2. For $i = 1$ to n

   Add X_i to the network

   select parents from $X_1, \ldots, X_{i-1}$ such that

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

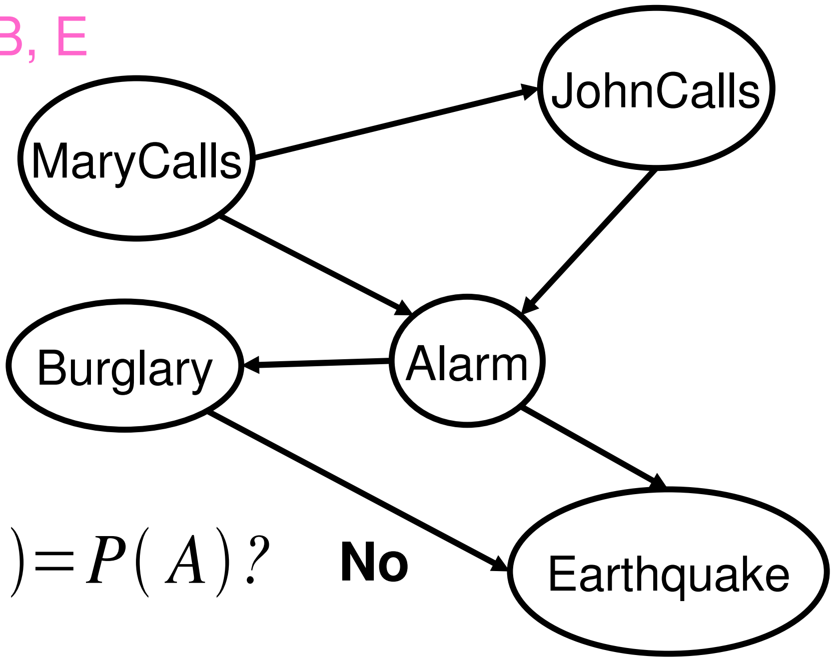
   This choice of parents guarantees the global semantics:

   $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1}) \quad \text{(chain rule)}$$

   $$= \prod_{i=1}^{n} P(X_i | Parents(X_i)) \quad \text{(by construction)}$$

Example

Suppose we choose the ordering M, J, A, B, E



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

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

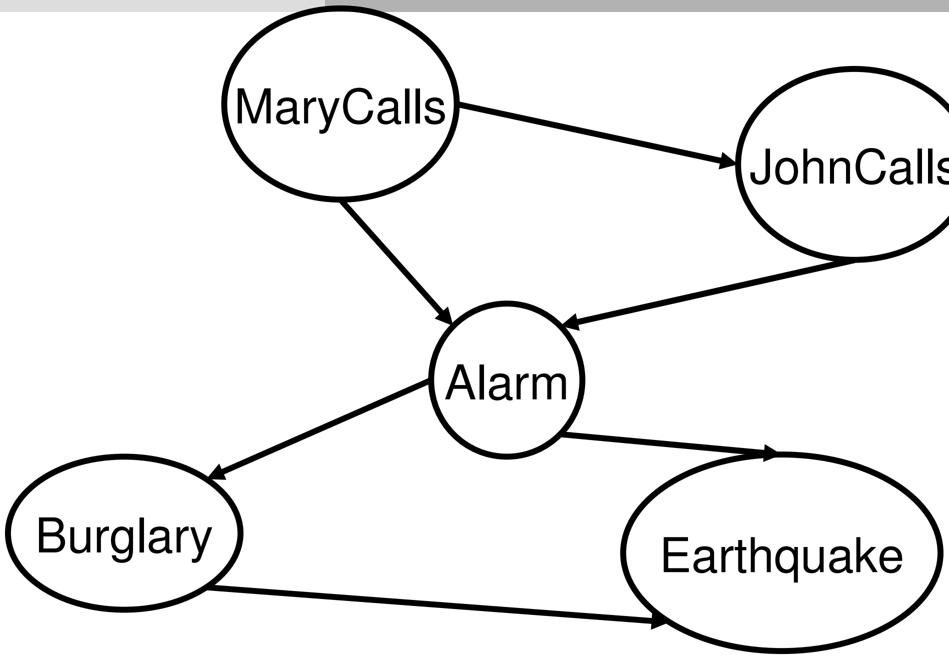
$P(B|A,J,M)=P(B|A)?$ **Yes**

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

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

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

Example continued:

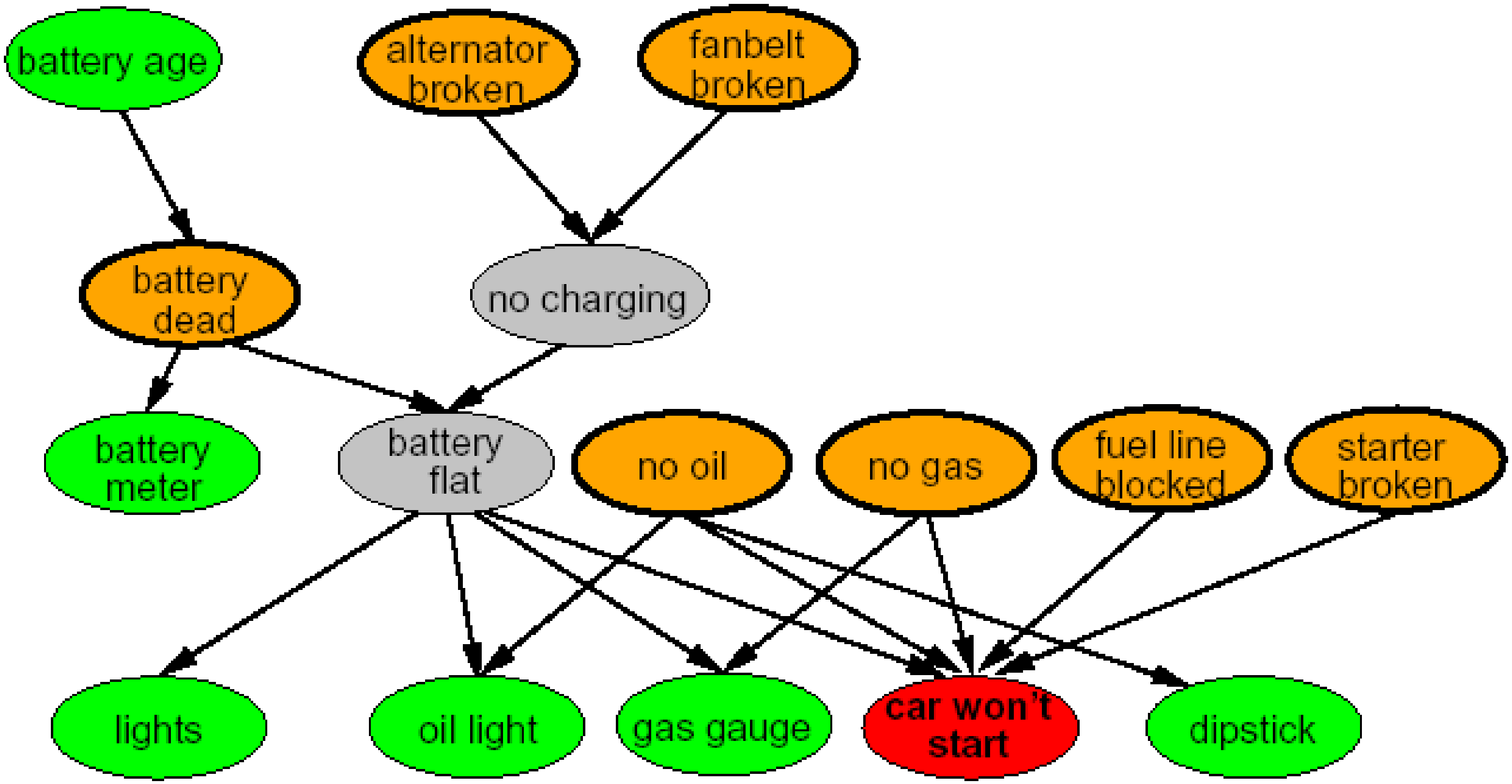


Deciding conditional independence is hard in noncausal directions
(Causal models and conditional independence seem hardwired for humans!)

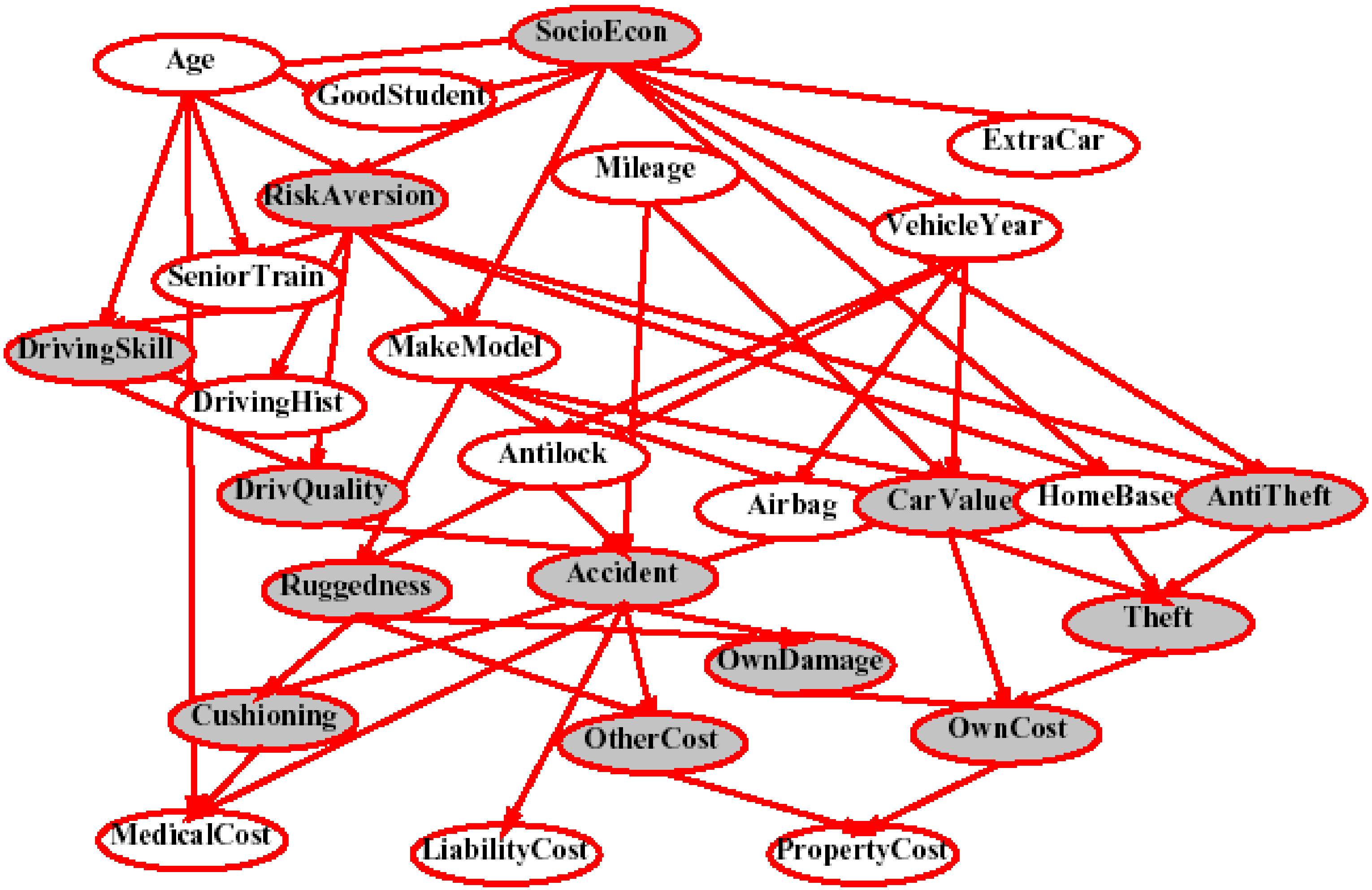
Assessing conditional probabilities is hard in noncausal directions and
Network is less compact: 1+2+4+3+4 = 13 numbers needed

Example: Car Diagnosis

Initial evidence: car won't start
Testable variables (green),
"broken, so fix it" variables (orange)
Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car Insurance



UNIVERSITÄT DES SAARLANDES

Compact conditional Distributions: **Deterministic Nodes**

CPT grows exponentially with number of parents
CPT becomes infinite with continuous-valued parent or child

Solution: canonical distributions that are defined more compactly

Deterministic nodes are the simplest case:

$$X = f(Parents(X))$$ for some function f

E.g., Boolean functions: “**NorthAmericans**”

NorthAmerican ⇔ Canadian ∨ US ∨ Mexican

E.g., numerical relationships among continuous variables: “**Lake Ontario**”

$$\frac{\partial Level}{\partial t} =$$ inflow + precipitation - outflow – evaporation


Compact conditional distributions: Noisy-Or Distributions

If: 1. Parent $U_1 \ldots U_k$ include all causes (possibly adding a leak node)

2. Independent failure probability q_i for each cause alone

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

Then: only k probabilities (those where the parent is true)

For Example:

fever if and only if *cold, flu, malaria* !

But: not always, it may be inhibited

Then, say:
P(~fever| cold, ~flu,~malaria) = 0.6
P(~fever| ~cold, flu, ~malaria) = 0.2
P(~fever| ~cold, ~flu, malaria) = 0.1

Number of parameters linear in number of parents

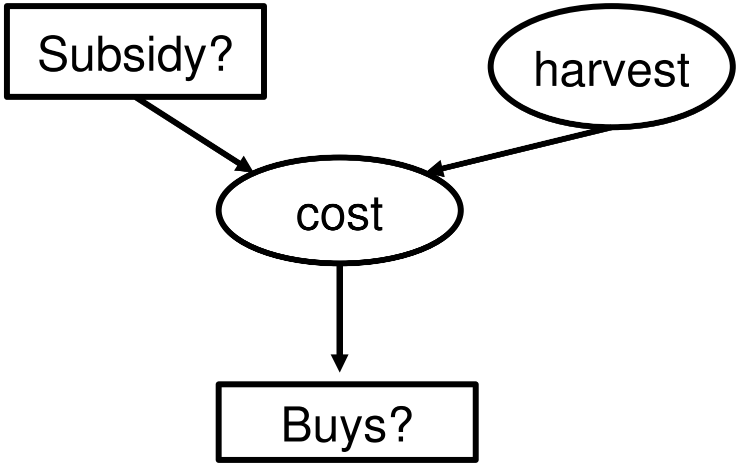
Compact conditional distributions: **Noisy-Or Distributions**

Cold	Flu	Malaria	P(Fever)	P(¬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 x 0.1
T	**F**	**F**	0.4	**0.6**
T	**F**	**T**	0.94	0.06 = 0.6 x 0.1
T	**T**	**F**	0.88	0.12 = 0.6 x 0.2
T	**T**	**T**	0.988	0.012 = 0.6 x 0.2 x 0.1

The probability is the product of the inhibition probabilities for each parent

Hybrid (discrete+continuous) Networks

Discrete (Subsidy? and Buys?); continuous (Harvest and Cost)



Option 1: discretization – possibly large errors, large CPTs

Option 2: finitely parameterized canonical families

1) Continuous variable, discrete+continuous parents (e.g., Cost)

2) Discrete variable, continuous parents (e.g., Buys?)

Summary

Bayes nets provide a natural representation for (causally induced) conditional independence

Topology + CPTs = compact representation of joint distribution

Generally easy for (non)experts to construct

Canonically distributions (e.g., noisy-OR) = compact representation of CPTs