Bayesian networks

Chapter 14.1–3

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:

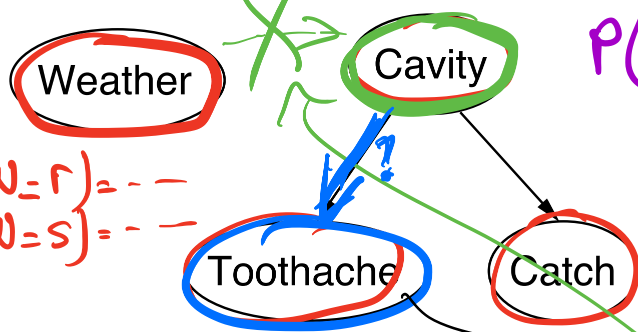
  $\mathbf{P}(X_i|Parents(X_i))$

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

$P(t \mid w, c, ct)$

Example

Topology of network encodes conditional independence assertions:



$P(T \mid \cancel{Catch}, Cavity)$

$P(W=r) = - - -$

$P(W=s) = - - -$

$P(Tooth=true)$

c	- —
$\neg c$	— —

$Weather$ is independent of the other variables

$Toothache$ and $Catch$ are conditionally independent given $Cavity$

$P(Cavity=true \mid Weather= \ldots) \overset{?}{=} P(Cavity=true)$

Answered Yes

$P(Toothache=true \mid Cavity) \overset{?}{=} P(Toothache=true)$

$P(\ \_\_\_\_\_\_\ \mid Weather)$ No $\overset{?}{=}$ - ''

Example

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?

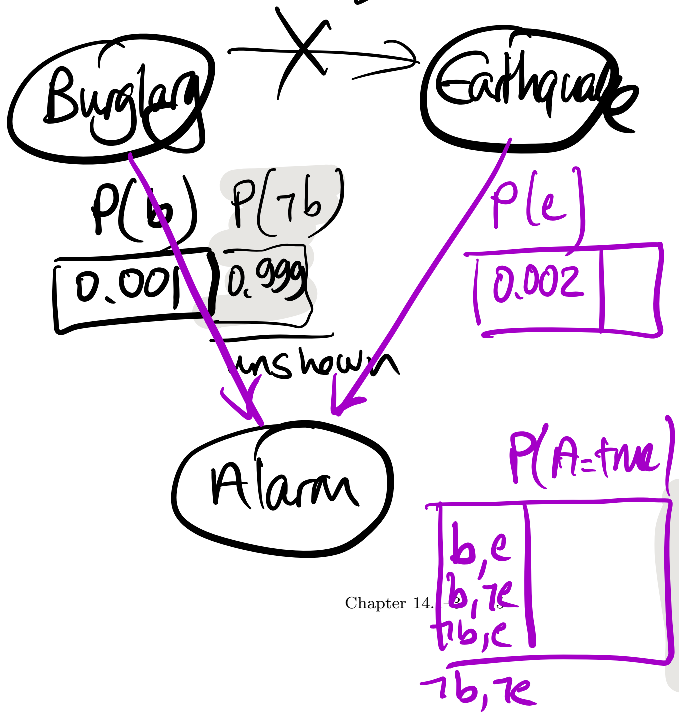
$P(E|B) \stackrel{?}{=} P(E)$ Yes

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$

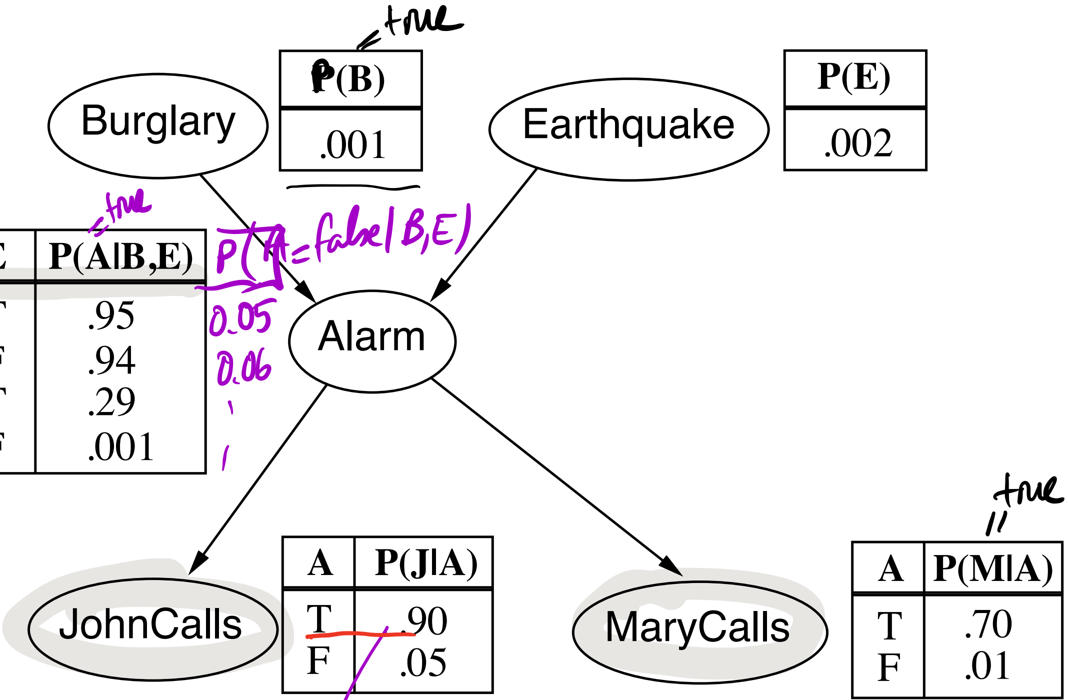
Network topology reflects "causal" knowledge:

- 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

Burglary
Earthquatee
JCalls
MCalls
Alarm



Example contd.



true

P(B)
.001

P(E)
.002

true

	B	E	P(A|B,E)	P(A=false|B,E)
b, e	T	T	.95	0.05
b, ¬e	T	F	.94	0.06
¬b, e	F	T	.29	
¬b, ¬e	F	F	.001	

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

true

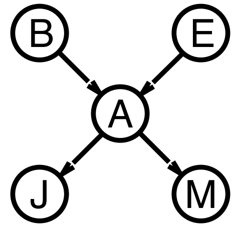
A	P(M|A)
T	.70
F	.01

$P(JC \mid a) = 0.90$

||

$P(JC \mid a, e) = 0.90$

Compactness



A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

Each row requires one number p for $X_i = true$
(the number for $X_i = false$ is just $1 - p$)

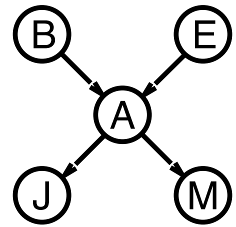
If each variable has no more than k parents,
the complete network requires $O(n \cdot 2^k)$ numbers

I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution

For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)

Global semantics

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



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

e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$: Prob. John calls, M " s t.i. an alarm.

$$= P(b) \times P(e|\cancel{b}) \times$$

$$\underline{P(a|e,b)} \times P(J|a,\cancel{e},\cancel{b}) \times P(M|a,\cancel{e},\cancel{b},\cancel{J})$$

by just successive chain rule

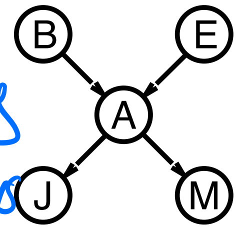
due to conditional indep. assumptions.

$$\ldots \; P(a|\underline{m,J})\,\underline{P(M|J)\,P(J)}$$

Joint Prob. $= \prod_{X_i=1}^{n} P(X_i | parents(X_i))$

Global semantics

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



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

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

$$= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$\approx 0.00063$$

summing thru the hidden vars

$P(b, J, m, a, e) +$
$P(b, J, m, a, \neg e) +$
$P(b, J, m, \neg a, e)$
$P(\quad, \neg a, \neg a)$

$$P(b | J, m) = \frac{P(b, J, m)}{P(J, m)} = \alpha \, P(b, J, m)$$

Quest of interest

go to the joint

$= \alpha \sum$

$$P(\text{lefthanded} \mid \text{male}) = 0.2$$

$$P(\quad \text{''} \quad \mid \text{female}) = 0.1$$

$$P(\text{male}) = 0.55$$

$$P(\text{female}) = 0.45$$

$$P(\text{lefthanded}) = ?$$

$$\|$$

$$P(\text{lefty}, \text{male}) + P(\text{lefty}, \text{female})$$

$$\|$$

$$P(\text{lefty} \mid \text{male}) \times P(\text{male})$$

$$P(\text{lefty} \mid \text{female}) P(\text{female})$$

Constructing Bayesian networks

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

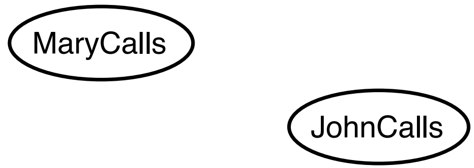
1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $$\mathbf{P}(X_i|Parents(X_i)) = \mathbf{P}(X_i|X_1,\, \ldots,\, X_{i-1})$$

This choice of parents guarantees the global semantics:

$$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1,\, \ldots,\, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|Parents(X_i)) \quad \text{(by construction)} \end{aligned}$$

Bad Ordering Example

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

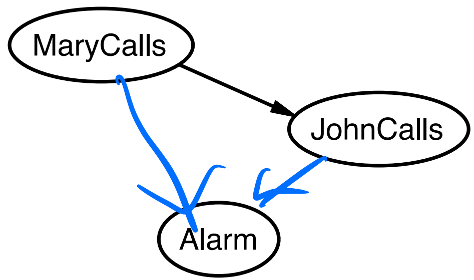


$P(J|M) = P(J)$?

Example

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

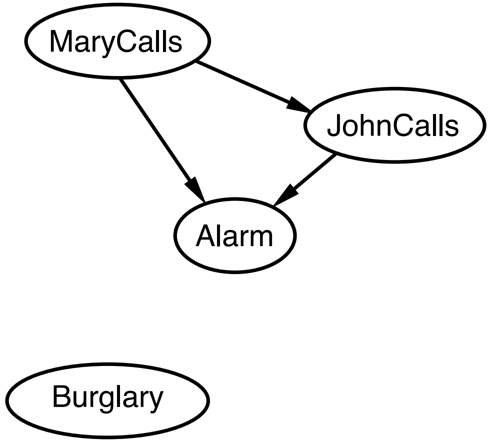
$P(\text{JCalls} \mid \text{MCalls})$



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

Example

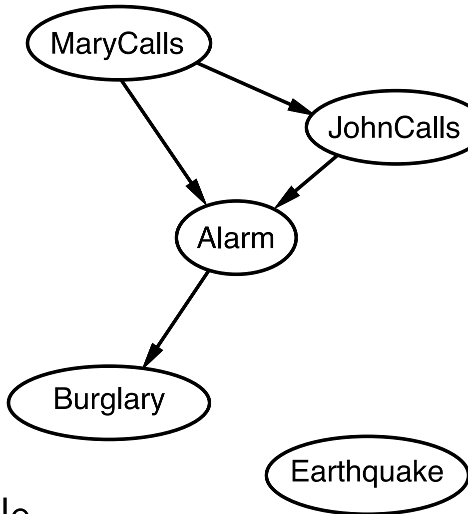
Suppose we choose the ordering M, J, A, B, E



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

Example

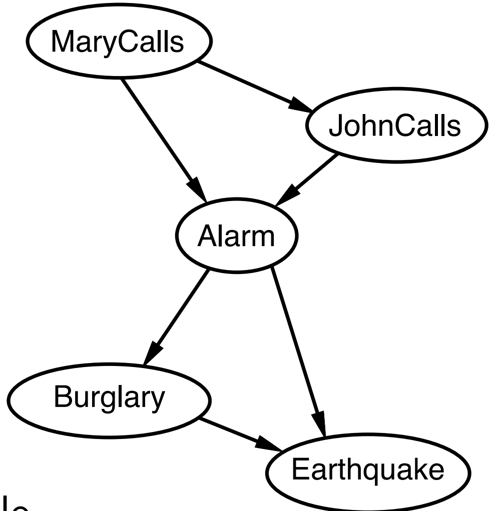
Suppose we choose the ordering M, J, A, B, E



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

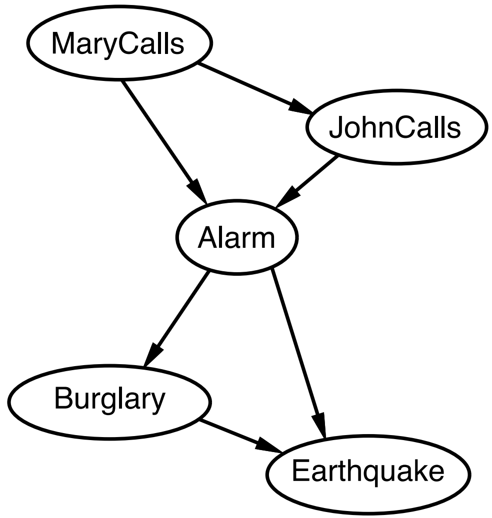
Example

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 contd.



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

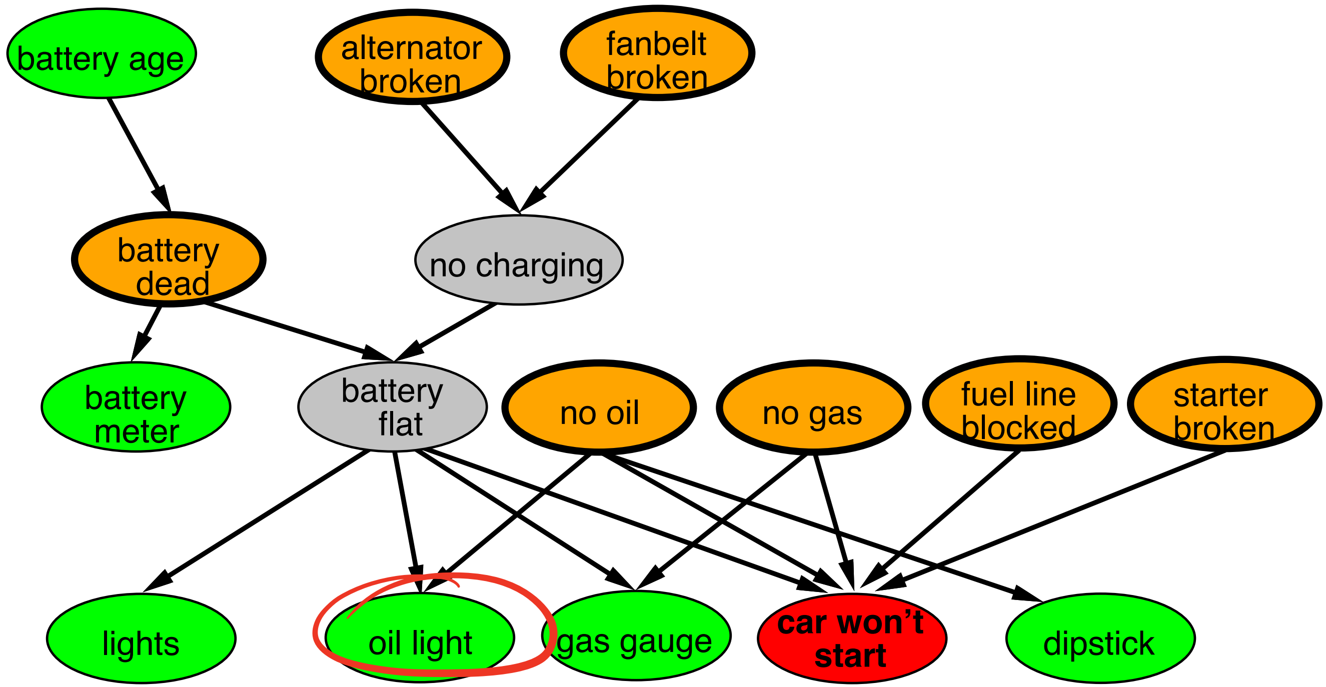
Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed

Example: Car diagnosis

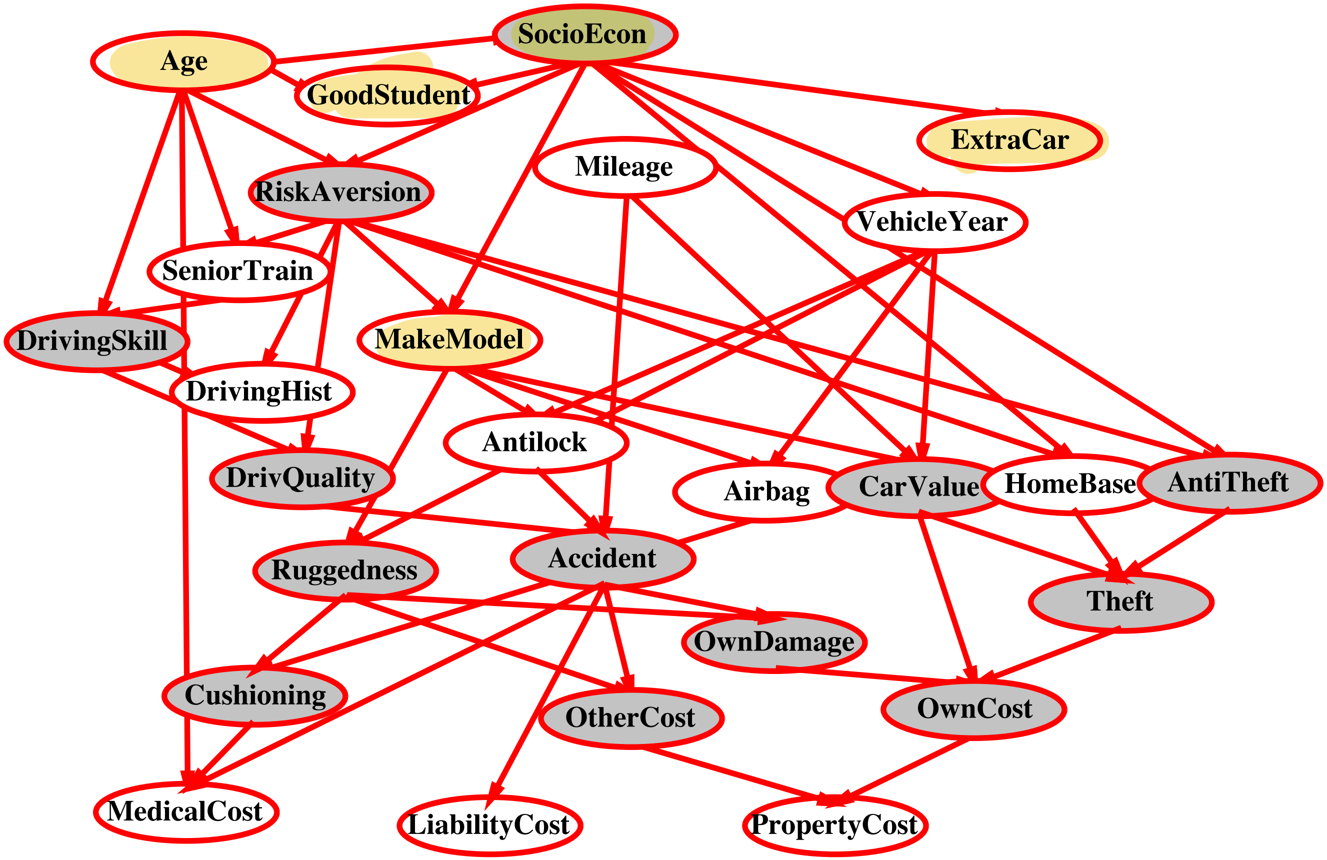
Initial evidence: car won't start

Testable variables (green), "broken, so fix it" variables (orange)

Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car insurance



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

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

Continuous variables $\Rightarrow$ parameterized distributions (e.g., linear Gaussian)