Bayesian networks

Chapter 14.1–4

Outline

◇ Syntax

◇ Semantics

◇ Exact inference by enumeration

◇ Exact inference by variable elimination

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

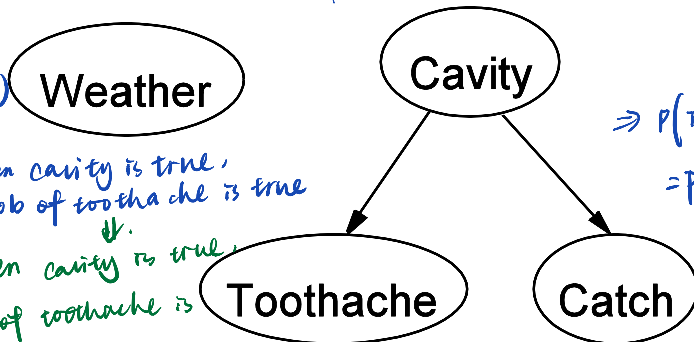
In the simplest case, conditional distribution represented as
a conditional probability table (CPT) giving the
distribution over X_i for each combination of parent values

Example 1

Topology of network encodes conditional independence assertions:

prior prob $P(\text{Cavity}) = \langle 0.1, 0.9 \rangle$

direct link from cavity to toothache



Cavity	P(toothach | Cavity)
T	0.8
F	0.1

→ given cavity is true, prob of toothache is true

⇓

given cavity is true, prob of toothache is false is 0.2

⇓

given cavity is false, prob of toothache is true 0.9.

$$\Rightarrow P(\text{Toothach}, \text{Catch}, \text{Cavity}) = P(\text{Toothach} \mid \text{Cavity}) \times P(\text{Catch} \mid \text{Cavity}) \times P(\text{Cavity})$$

Weather is independent of the other variables

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

sum over hidden variable

$$\text{eg. } P(\text{cavity} \mid \text{toothache}) = \frac{P(\text{cavity}, \text{toothache})}{P(\text{toothache})} = \frac{\sum_{\text{Catch}} P(\text{cavity}, \text{toothache}, \text{Catch})}{\sum_{\text{cavity}} \sum_{\text{catch}} P(\text{toothache}, \text{Catch}, \text{Cavity})}$$

$$= \frac{P(\text{toothache} \mid \text{cavity}) \cdot P(\text{cavity})}{\sum_{\text{cavity}} P(\text{toothache} \mid \text{Cavity})\, P(\text{Cavity})}$$

$$= \frac{P(\text{toothache} \mid \text{cavity}) \cdot P(\text{cavity})}{P(\text{toothache} \mid \text{cavity}) \cdot P(\text{cavity}) + P(\text{toothache} \mid \neg\text{cav})\, P(\neg\text{cav})}$$

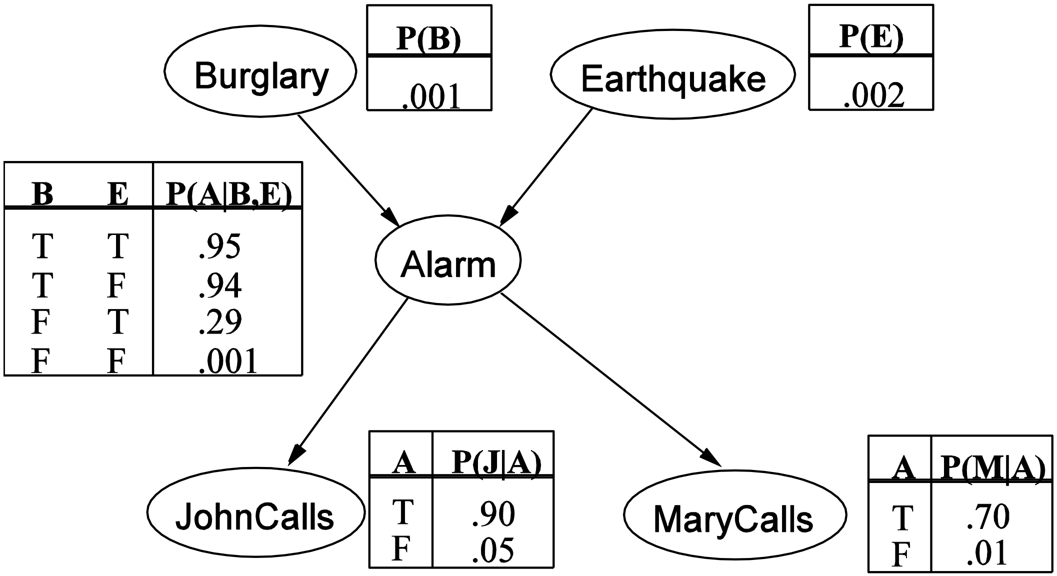
Example 2

Scenario: 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?

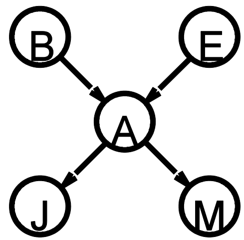
Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$

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 contd.



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 $1 - p$)

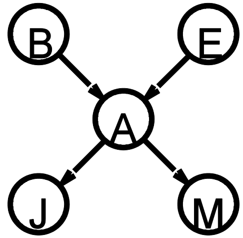
If each variable has no more than k parents, n. variables
the complete network requires $O(n \cdot 2^k)$ numbers

if k is small ?
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. $2^5 - 1 = 31$)

prob should be summed to 1

Global semantics



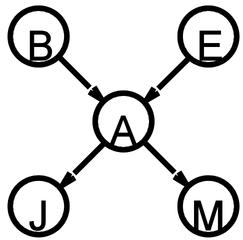
Global semantics defines the full joint distribution as the product of the local conditional distributions:

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

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

$$=$$

Global semantics



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

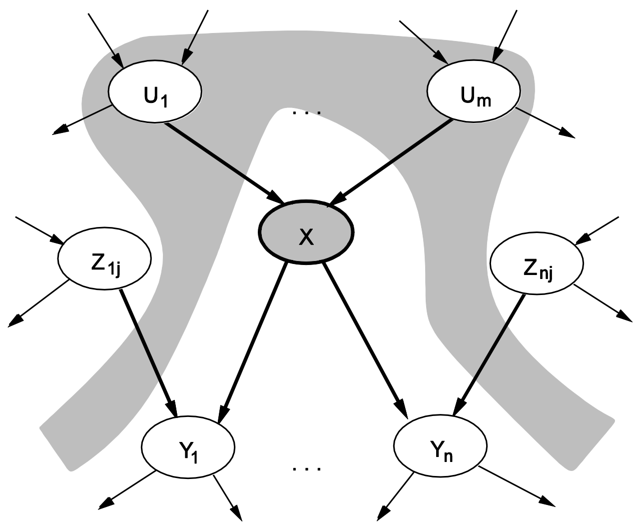
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\mathsf{parents}(X_i))$$

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

$$\begin{aligned} &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$

Local semantics

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



Constructing Bayesian networks

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

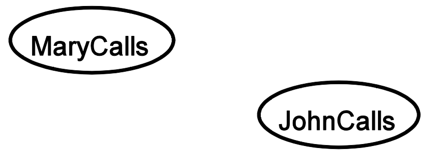
1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $$\mathbf{P}(X_i|\text{Parents}(X_i)) = \mathbf{P}(X_i|X_1,\ \ldots,\ X_{i-1})$$

This choice of parents guarantees the global semantics:

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

Example

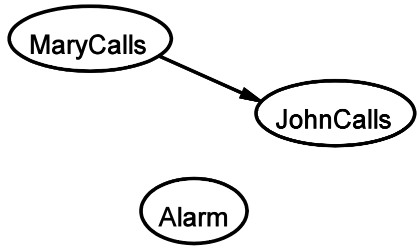
Suppose we choose the ordering M, J, A, B, E



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

Example

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

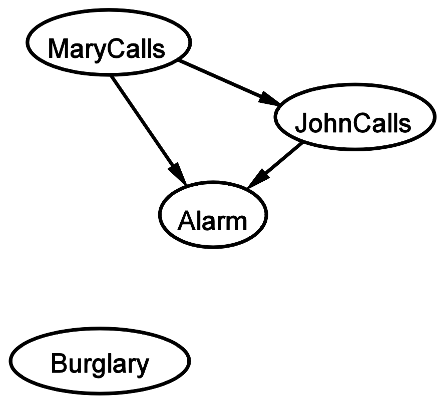


correlation between J & M

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

Example

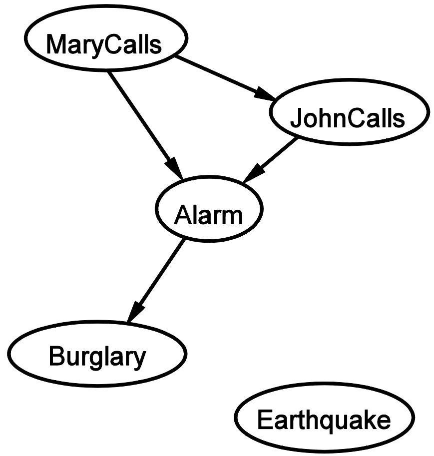
Suppose we choose the ordering M, J, A, B, E



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

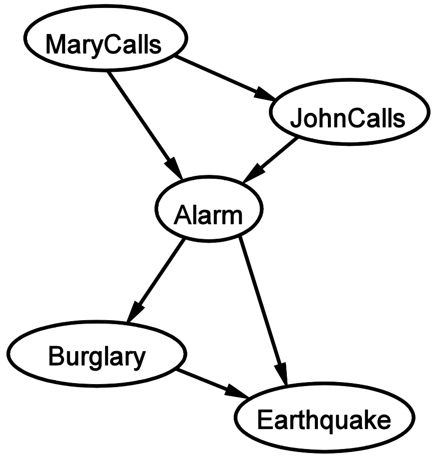
Suppose we choose the ordering M, J, A, B, E



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

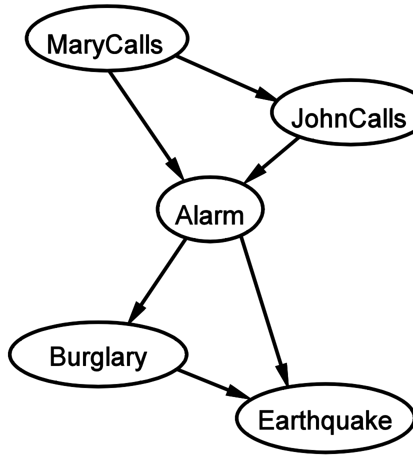
Example

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 (symptoms $\rightarrow$ causes)

MaryCall & John Call $\rightarrow$ symptom

Causal models (causes $\rightarrow$ symptoms) and conditional independence seem easier for humans!

Assessing conditional probabilities is hard in noncausal directions

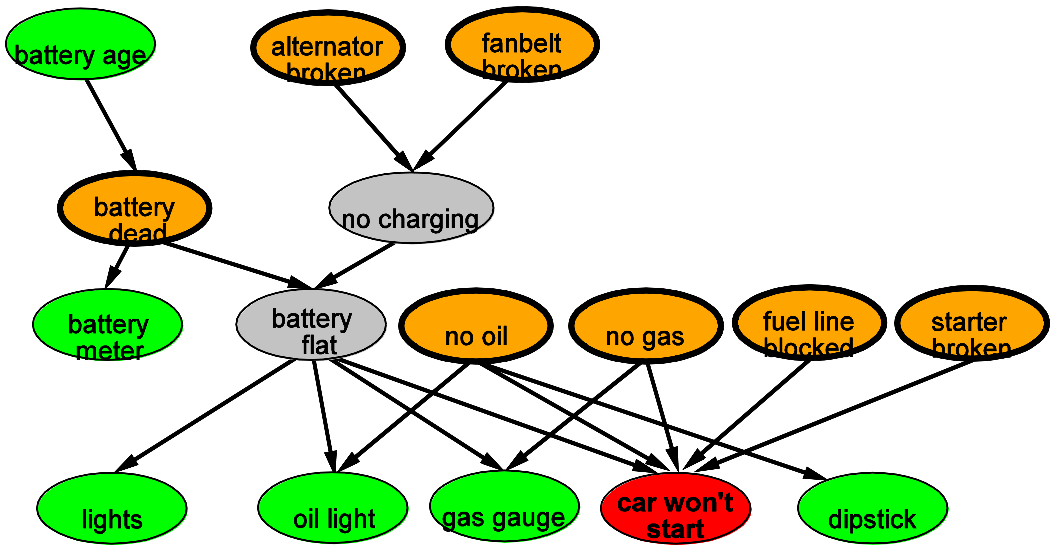
Network is less compact: $1+2+4+2+4=13$ numbers needed

Example: Car diagnosis

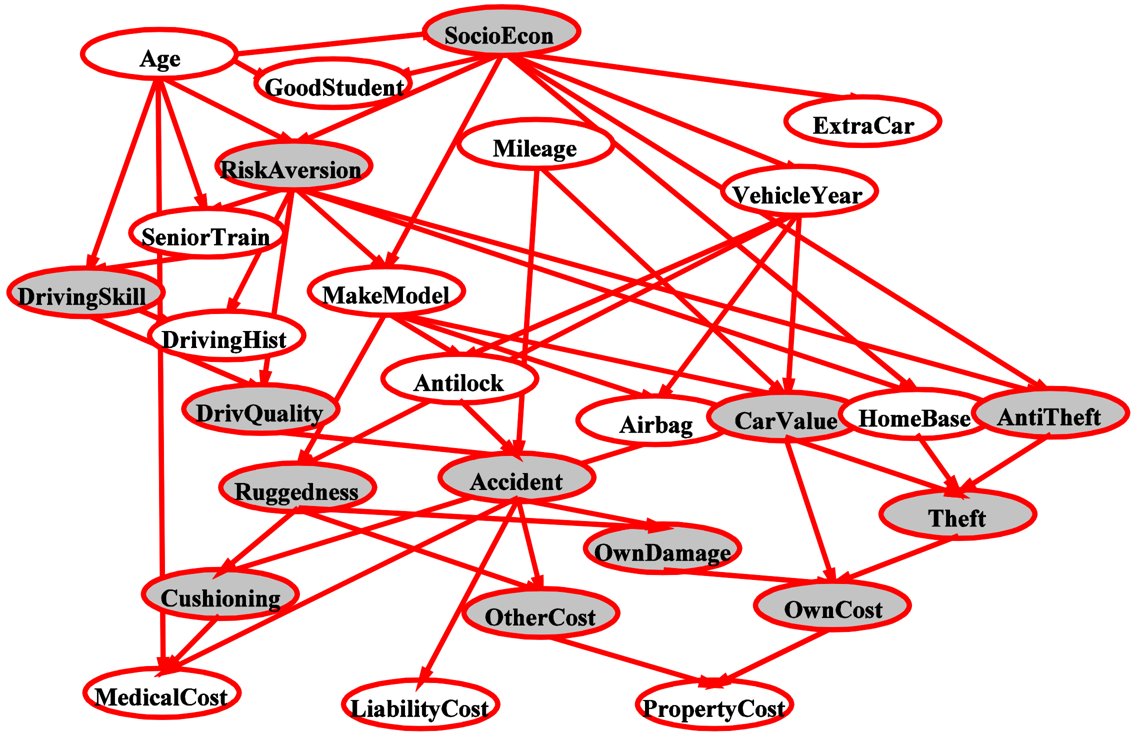
Initial evidence: car won't start

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

Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car insurance



Inference tasks

Simple queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=E)$
e.g., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

Conjunctive queries: $\mathbf{P}(X_i, X_j|\mathbf{E}=E) = \mathbf{P}(X_i|\mathbf{E}=E)\mathbf{P}(X_j|X_i, \mathbf{E}=E)$

probability is estimated.

Value of information: which evidence to seek next?

Sensitivity analysis: which probability values are most critical?

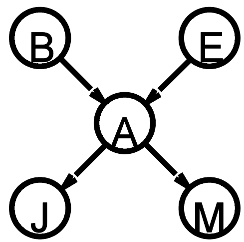
Explanation: why do I need a new starter motor?

We focus on **simple** and **conjunctive** queries

Inference by enumeration

Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

Simple query on the burglary network:



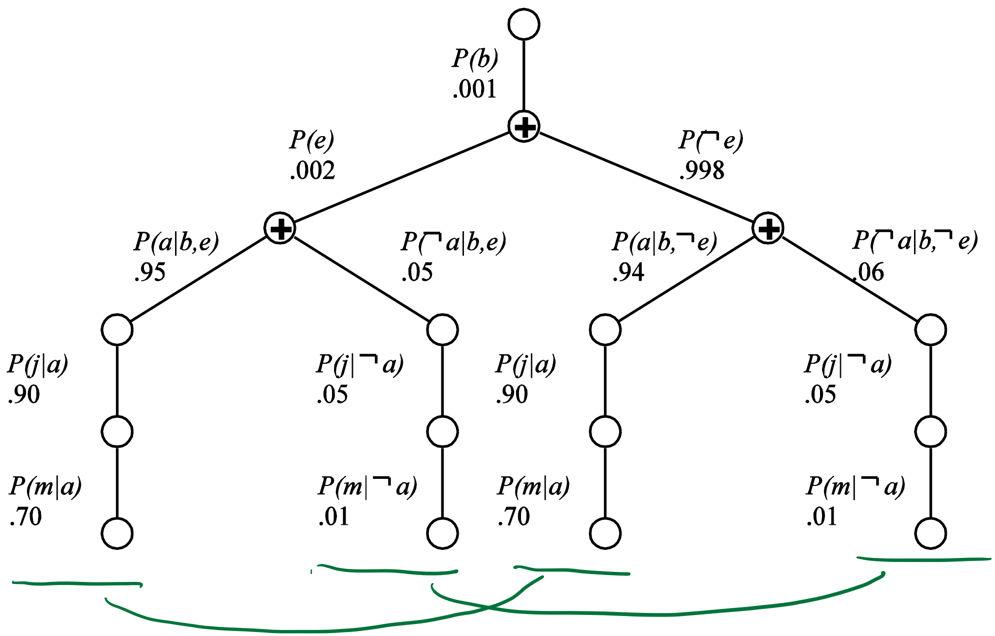
$\mathbf{P}(B|j,m)$
$= \mathbf{P}(B,j,m)/P(j,m)$
$= \alpha\mathbf{P}(B,j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$

Rewrite full joint entries using product of CPT entries:

$\mathbf{P}(B|j,m)$ doesn't depend on e, a
$= \alpha \sum_e \sum_a \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$ move some terms out of Σ
$= \alpha\mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B,e)P(j|a)P(m|a)$

Recursive depth-first enumeration: $O(n)$ space, $O(d^n)$ time
possible value

Evaluation tree



Enumeration is inefficient: repeated computation

e.g., computes $P(j|a)P(m|a)$ for each value of e

can use memorisation for some nodes

Inference by variable elimination

Variable elimination: carry out summations right-to-left,
storing intermediate results (factors) to avoid recomputation

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

Variable elimination: Basic operations

Summing out a variable from a product of factors:
move any constant factors outside the summation
add up submatrices in pointwise product of remaining factors

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

assuming $f_1, \ldots, f_i$ do not depend on X

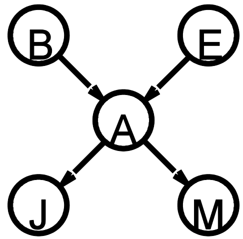
Pointwise product of factors f_1 and f_2:

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

e.g., $f_1(a, b) \times f_2(b, c) = f(a, b, c)$

Irrelevant variables

Consider the query $P(JohnCalls|Burglary\!=\!true)$



$$P(J|b) = \alpha P(b) \sum_e P(e) \sum_a P(a|b,e) P(J|a) \sum_m P(m|a)$$

Sum over m is identically 1; M is **irrelevant** to the query

Thm 1: Y is irrelevant unless $Y \in Ancestors(\{X\} \cup \mathcal{E}_)$

Here, $X = JohnCalls$, $\mathcal{E}_= \{Burglary\}$, and
$Ancestors(\{X\} \cup \mathcal{E}_) = \{Alarm, Earthquake\}$
so $MaryCalls$ is irrelevant

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

Exact inference by enumeration

Exact inference by variable elimination

Examples of skills expected:

- ♢ Formulate a belief network for a given problem domain
- ♢ Derive expression for joint probability distribution for given belief network
- ♢ Use inference by enumeration to answer a query about simple or conjunctive queries on a given belief network