Artificial Intelligence

Hai Thi Tuyet Nguyen

Outline

CHAPTER 8: PROBABILISTIC REASONING

8.1 Representing Knowledge In An Uncertain Domain

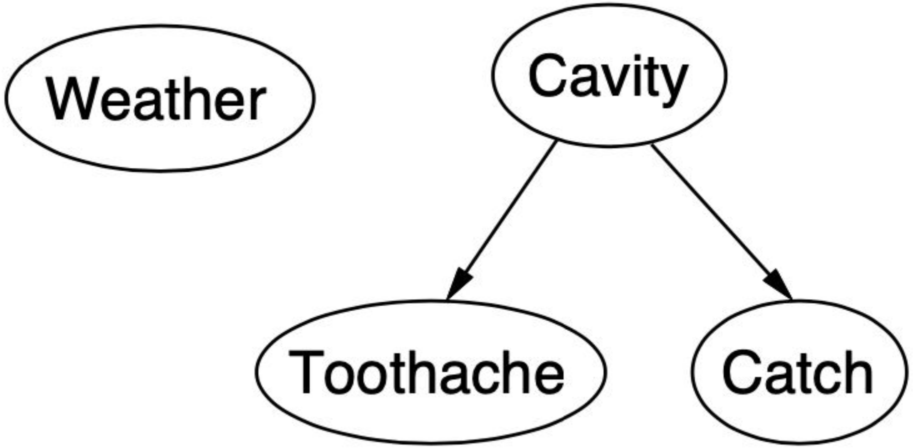
8.1 Representing Knowledge In An Uncertain Domain

- A simple, graphical notation for conditional independence assertions
- 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

8.1 Example

Topology of network encodes conditional independence assertions:



P(Cavity | toothache $\wedge$ catch) = α **P**(toothache | Cavity) **P**(catch | Cavity) **P**(Cavity)

Weather is independent of the other variables

Toothache and Catch are conditionally independent given Cavity

8.1 Example

Exercise: You have a new burglar alarm installed at home. It is fairly reliable at detecting a burglary, but also responds on occasion to minor earthquakes. You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm. John nearly always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too. Mary, on the other hand, likes rather loud music and often misses the alarm altogether. Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

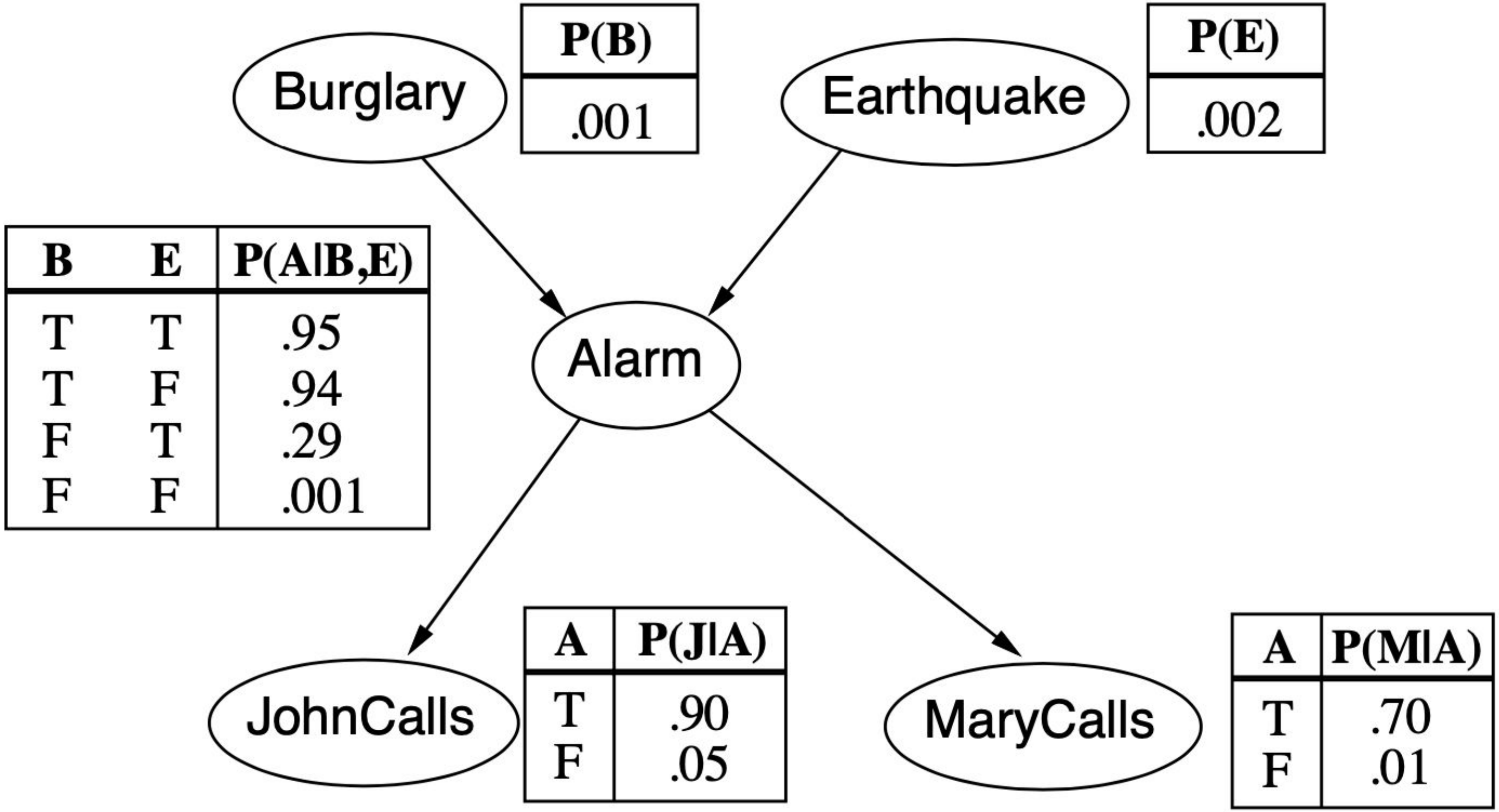
8.1 Example

Exercise: You have a new burglar alarm installed at home. It is fairly reliable at detecting a burglary, but also responds on occasion to minor earthquakes. You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm. John nearly always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too. Mary, on the other hand, likes rather loud music and often misses the alarm altogether. Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

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

8.1 Example



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

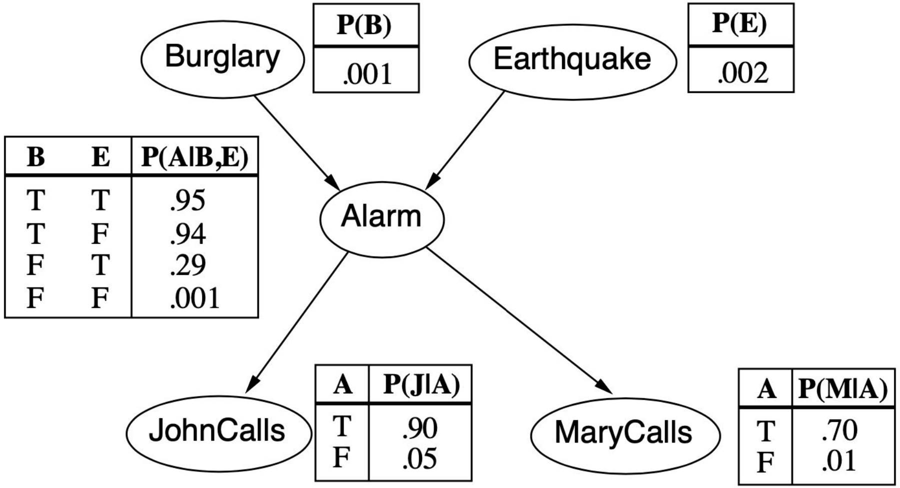
8.2 The Semantics Of Bayesian Networks

8.2 Global semantics

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

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

P(j ∧ m ∧ a ∧ ¬b ∧ ¬e)



8.2 Global semantics

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

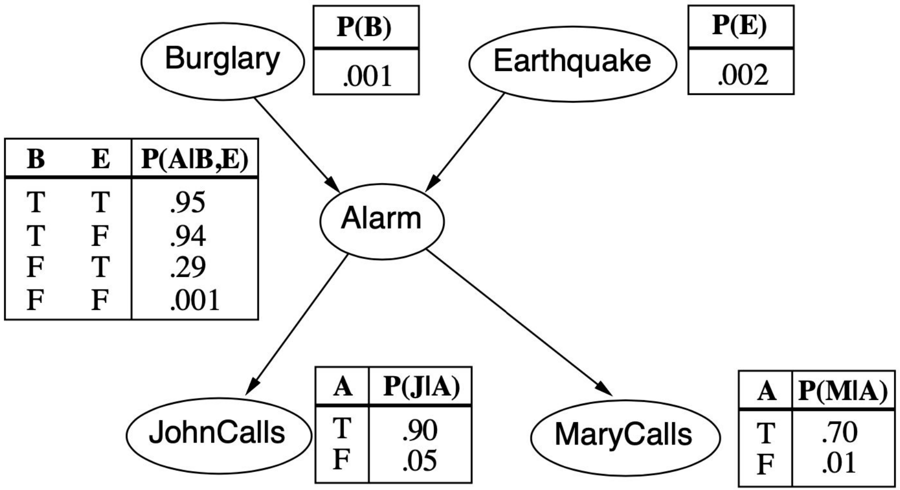
$$P(x_1, \ldots, x_n) = \Pi_{i=1}^{n} P(x_i | parents(X_i))$$

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

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

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$

≈ 0.00063

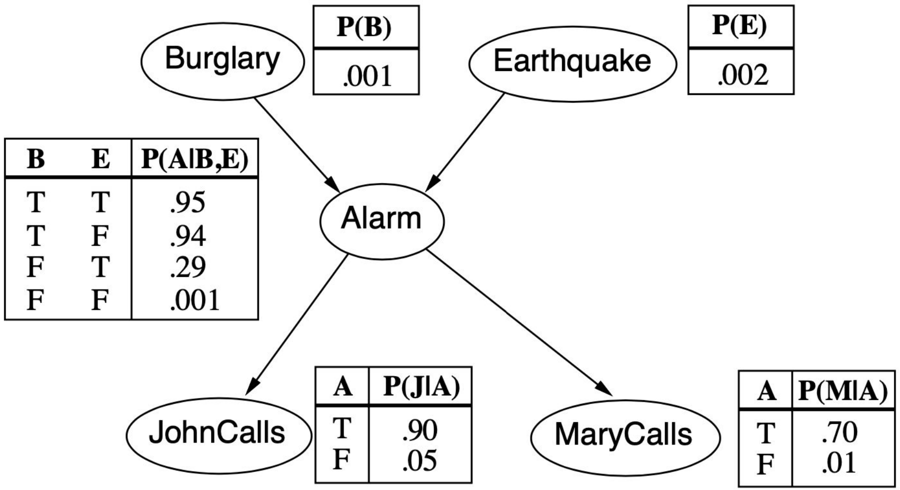


8.2 Global semantics

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

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

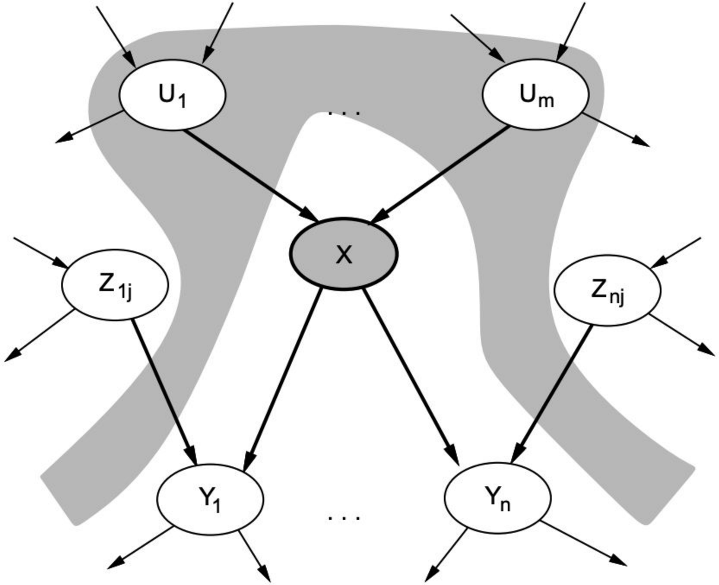
$P(j \wedge \neg m \wedge \neg a \wedge b \wedge e)$

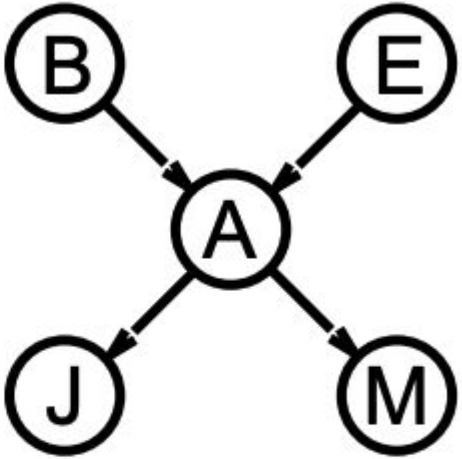


8.2 Local semantics

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

For example, JohnCalls is independent of Burglary, Earthquake, and MaryCalls given the value of Alarm.

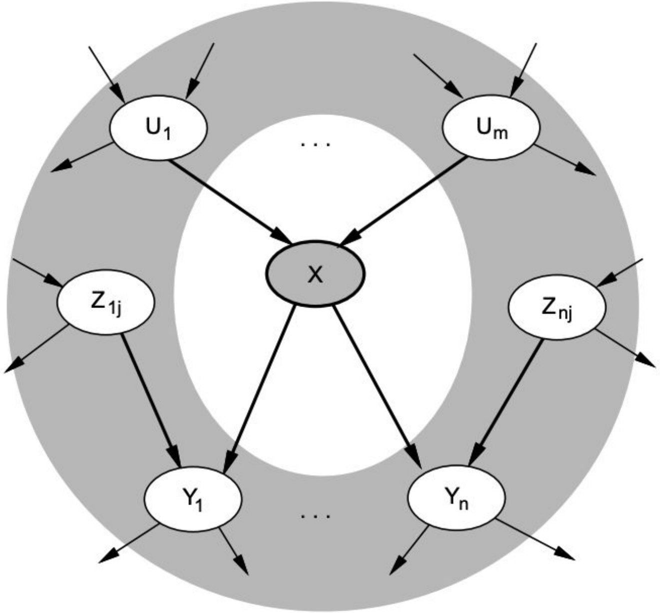


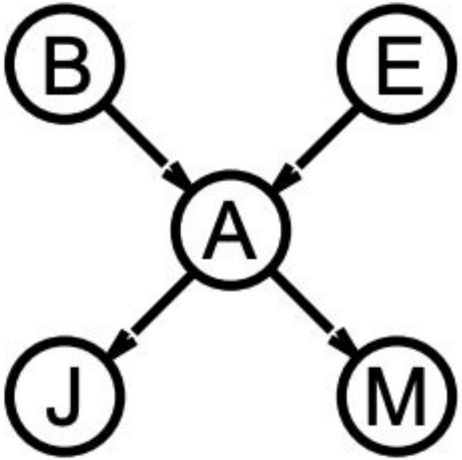


8.2 Markov blanket

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

For example, Burglary is independent of JohnCalls and MaryCalls, given Alarm and Earthquake.





8.2 Constructing Bayesian networks

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i=1$ to n do

    add X_i to the network

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

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

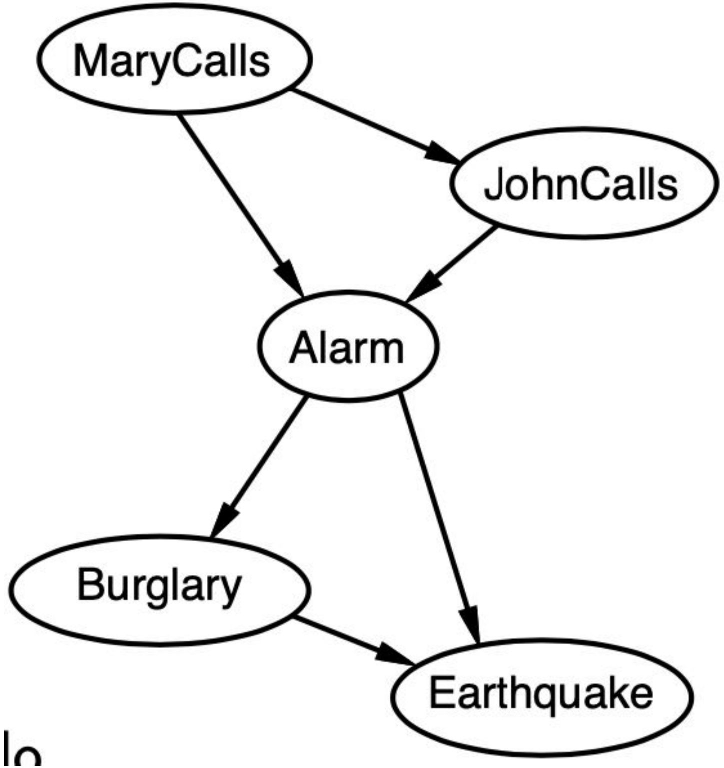
This choice of parents guarantees the global semantics:

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

8.2 Example

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

- Adding MaryCalls: No parents.
- Adding JohnCalls: If Mary calls, which would make it more likely that John calls.
  => JohnCalls needs MaryCalls as a parent.
- Adding Alarm: Clearly, if both call, it is more likely that the alarm has gone off if just one or neither calls
  => MaryCalls and JohnCalls as parents.

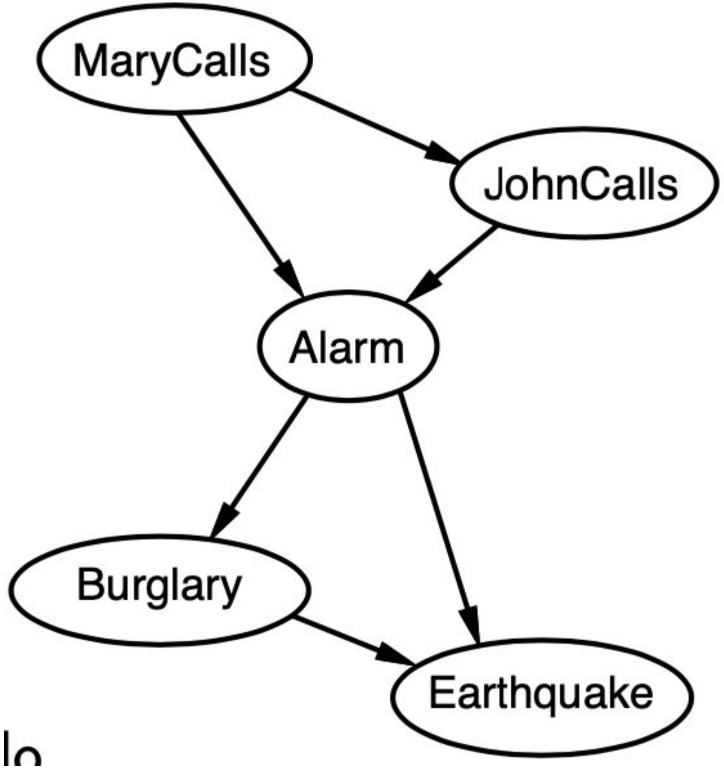


8.2 Example

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

- Adding Burglary: If we know the alarm state, then the call from John or Mary might give us information about our phone ringing, but not about burglary:
  P(Burglary | Alarm, JohnCalls, MaryCalls)
  = P(Burglary | Alarm) .
  => Alarm as parent.

- Adding Earthquake: If the alarm is on, it is more likely that there has been an earthquake.
  But if we know that there has been a burglary, then that explains the alarm
  => Alarm and Burglary as parents.



8.2 Example

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

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

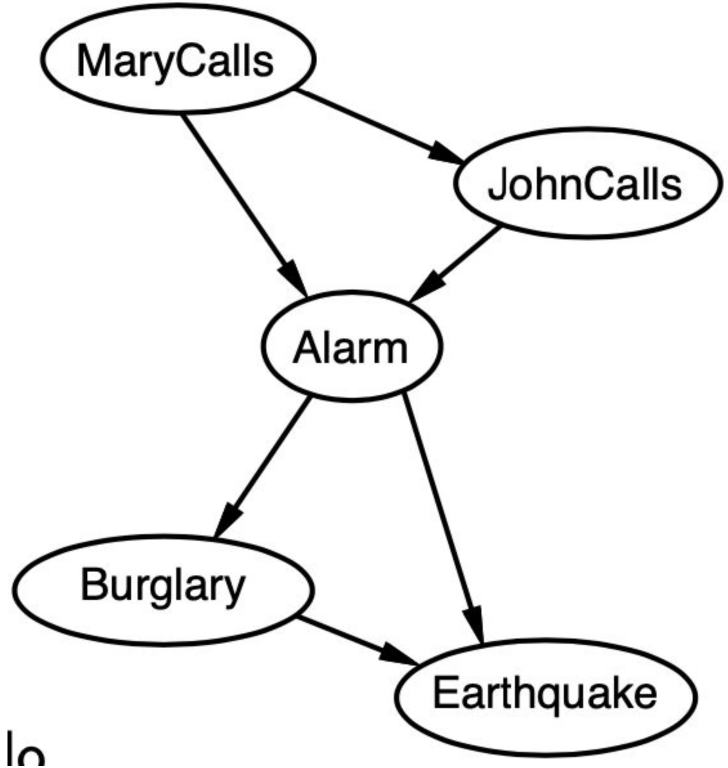
$P(A|J, M) = P(A)$? No

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



8.2 Example

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

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

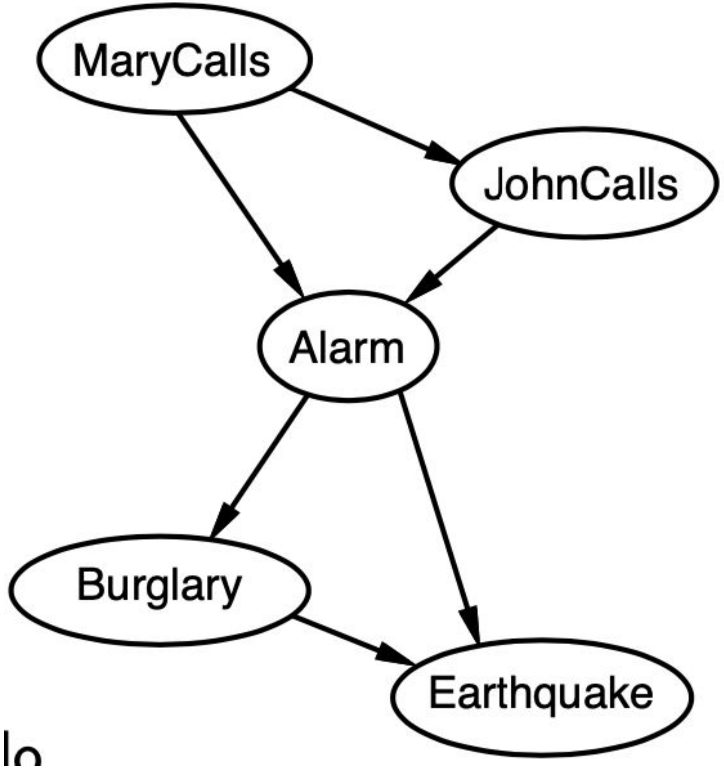
$P(A|J, M) = P(A)$?

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

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

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

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

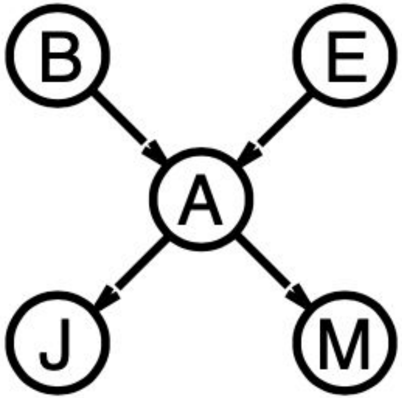


8.3 Exact Inference In Bayesian Networks

8.3 Inference by enumeration

- Simple queries: compute the posterior probability distribution $\mathbf{P}(X_i|E = e)$
  e.g., P(NoGas=true|Gauge=empty, Lights=on, Starts=false)
- A query can be answered using a Bayesian network by computing sums of products of conditional probabilities from the network
  Compute **P**(Burglary|JohnCalls=true,MaryCalls=true),
  the hidden variables for this query are Earthquake and Alarm

$$\mathbf{P}(B|j, m) = \alpha\Sigma_e\, \Sigma_a\, \mathbf{P}(B, j, m, e, a)$$



For Burglary = true

$$P(b|j, m)= \alpha\Sigma_e\, \Sigma_a\, P(b)\, P(e)\, P(a|b,e)\, P(j|a)\, P(m|a)$$
$$= \alpha\, P(b)\, \Sigma_e P(e)\, \Sigma_a P(a|b,e)\, P(j|a)\, P(m|a)$$

$P(b, j, m, e, a)$

$P(b)\, P(e)\, P(a|b, e)\, P\,(j|a)\, P\,(m|a)$

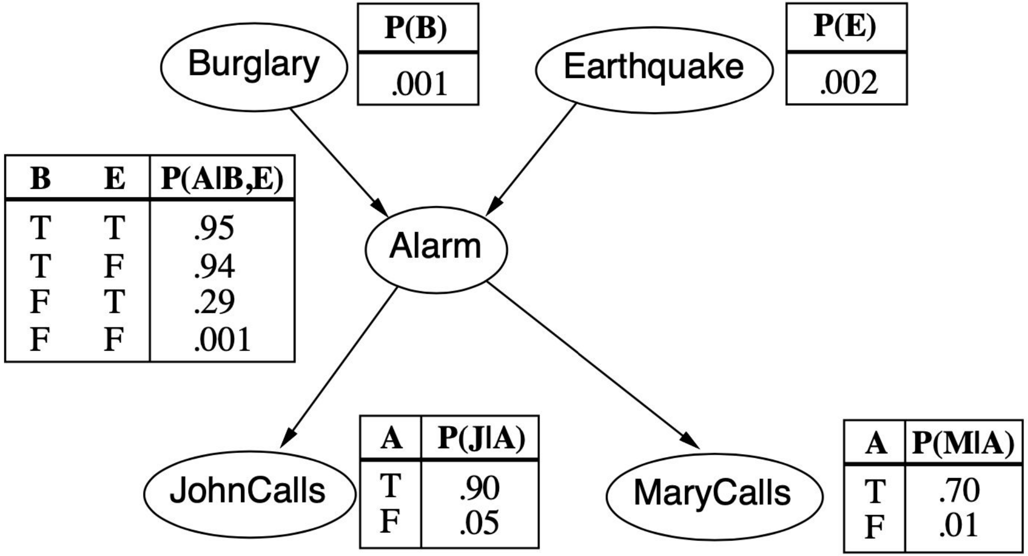
$\mathbf{P}(B|j, m)$

$= \alpha[\ \mathbf{P}(B,j, m, e, a)+$

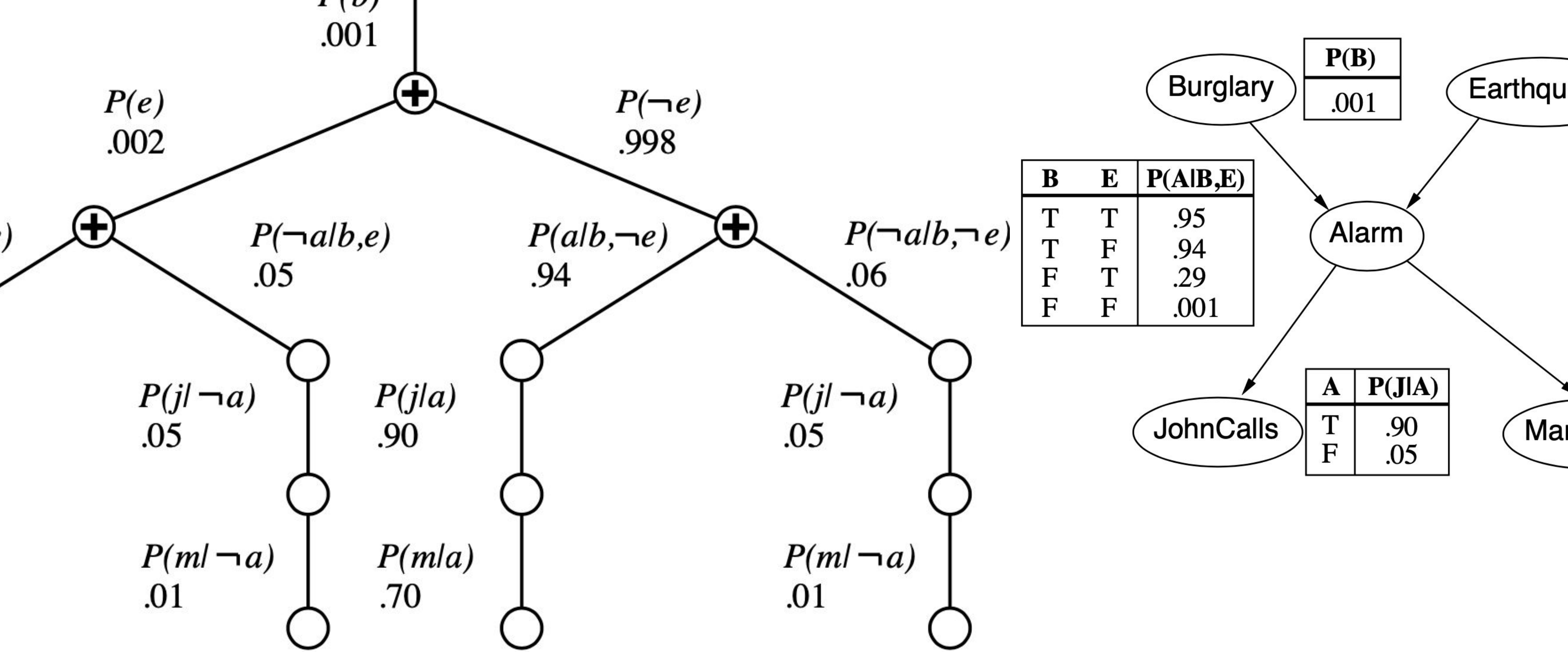
$\quad \mathbf{P}(B,j, m, \neg e, a)+$

$\quad \mathbf{P}(B,j, m, e, \neg a)+$

$\quad \mathbf{P}(B,j, m, \neg e, \neg a)]$



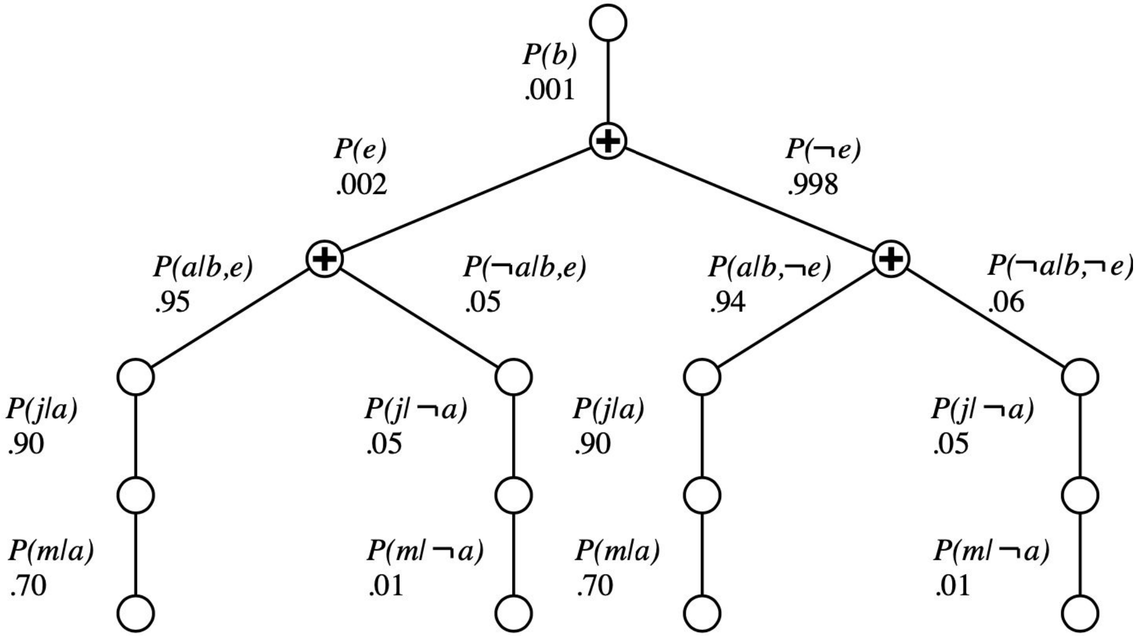
$\mathbf{P}(B,j, m, e, a)$: $\langle P(b, j, m, e, a), P(\neg b, j, m, e, a)\rangle$



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

- The evaluation proceeds top down, multiplying values along each path and summing at the “+” nodes.

8.3 Inference by enumeration



- Enumeration is inefficient: repeated computation
  e.g., computes P(j|a)P(m|a) for each value of e.

8.3 Inference by variable elimination

- Variable elimination:
  - carry out summations right-to-left,
  - storing intermediate results (factors) to avoid recomputation
- Notes:
  - each factor is a matrix indexed by the values of its argument variables.
  - "×" operator is not ordinary matrix multiplication but instead the pointwise product operation

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

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

8.3 Inference by variable elimination

$$\mathbf{f}_4(A) = \begin{pmatrix} P(j \mid a) \\ P(j \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix} \qquad \mathbf{f}_5(A) = \begin{pmatrix} P(m \mid a) \\ P(m \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}$$

First, we sum out A from the product of $\mathbf{f}_3$, $\mathbf{f}_4$, and $\mathbf{f}_5$. This gives us a new 2×2 factor $\mathbf{f}_6$(B, E)

$$\begin{aligned} \mathbf{f}_6(B, E) &= \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \\ &= (\mathbf{f}_3(a, B, E) \times \mathbf{f}_4(a) \times \mathbf{f}_5(a)) + (\mathbf{f}_3(\neg a, B, E) \times \mathbf{f}_4(\neg a) \times \mathbf{f}_5(\neg a)) . \end{aligned}$$

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E) .$$

8.3 Inference by variable elimination

Next, we sum out E from the product of $\mathbf{f}_2$ and $\mathbf{f}_6$:

$$\begin{aligned}\mathbf{f}_7(B) &= \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E) \\ &= \mathbf{f}_2(e) \times \mathbf{f}_6(B, e) + \mathbf{f}_2(\neg e) \times \mathbf{f}_6(B, \neg e)\ .\end{aligned}$$

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{f}_1(B) \times \mathbf{f}_7(B)$$

8.3 Inference by variable elimination

- The pointwise product of two factors $\mathbf{f_1}$ and $\mathbf{f_2}$ yields a new factor f whose variables are the union of the variables in $\mathbf{f_1}$ and $\mathbf{f_2}$
- Suppose the two factors have variables $Y_1, \ldots, Y_k$ in common.
  Then we have
  $\mathbf{f}(X_1 \ldots X_j, Y_1 \ldots Y_k, Z_1 \ldots Z_l) = \mathbf{f_1}(X_1 \ldots X_j, Y_1 \ldots Y_k)\ \mathbf{f_2}(Y_1 \ldots Y_k, Z_1 \ldots Z_l)$
- E.g., $\mathbf{f_1}(a, b) \times \mathbf{f_2}(b, c) = \mathbf{f}(a, b, c)$

8.3 Inference by varial

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

$$= \begin{pmatrix} .06 & .24 \\ .42 & .28 \end{pmatrix} + \begin{pmatrix} .18 & .72 \\ .06 & .04 \end{pmatrix} = ($$

- E.g., $\mathbf{f}_1(a, b) \times \mathbf{f}_2(b, c) = \mathbf{f}(a, b, c)$

A	B	$\mathbf{f}_1(A, B)$	B	C	$\mathbf{f}_2(B, C)$	A	B	C	$\mathbf{f}_3(A, B, C)$
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	$.1 \times .4 = .04$

8.3 Inference by variable elimination

- Summing out a variable from a product of factors:
    - add up submatrices in pointwise product of remaining factors
- For example, to sum out A from $f_3(A, B, C)$, we write

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

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

Inference by variable elimination

P(Burglary | JohnCalls = true, MaryCalls = true) ?