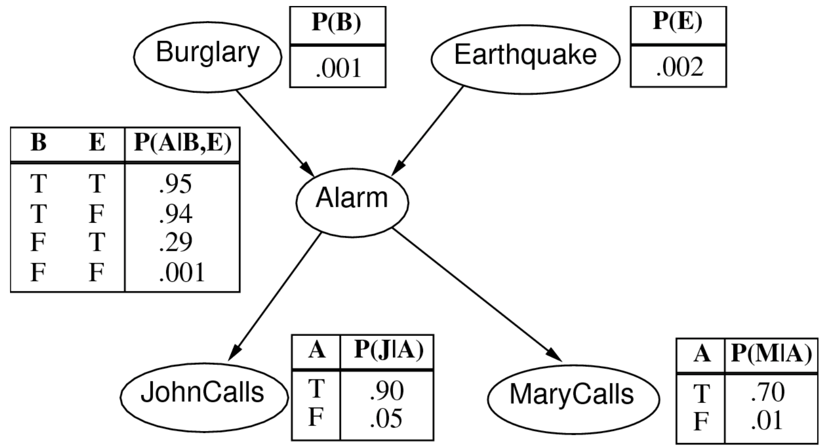
Exact Inference Methods

3007/7059 Artificial Intelligence

School of Computer Science
The University of Adelaide

An inference problem...

I'm at work, neighbour John called to say my alarm is ringing, but neighbour Mary didn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?



Recall the global semantics of a Bayesian Network

A Bayesian Network encodes our knowledge on how the variables interact. This is specified in terms of conditional independence assertions.

The global semantics of a network define a joint distribution of all variables as the product of local conditional distributions.

The joint distribution defined by a Bayesian Network with variables $X_1, \ldots, X_n$ is:

$$\begin{aligned} P(X_1, \ldots, X_n) &= P(X_1|Parents(X_1)) \times P(X_2|Parents(X_2)) \\ &\quad \times \cdots \times P(X_n|Parents(X_n)) \\ &= \prod_{i=1}^{n} P(X_i|Parents(X_i)) \end{aligned}$$

where $Parents(X_i)$ are parents of X_i as specified by the particular Bayesian Network.

Performing inference on Bayesian Networks

For the burglar-alarm problem, we wish to compute the probability that there is a burglar ($B = true$ or b) given the evidences John called ($J = true$ or j) but Mary didn't ($M = false$ or $\neg m$), i.e.,

$$P(b|j, \neg m)$$

Recall the general rule of statistical inference:

$$P(X|e) = \alpha \sum_{\forall Y} P(X, e, Y)$$

where X is the query variable, e the observed values for the evidence variables, and Y the unobserved variables. As usual α is a normalisation constant that we solve for at the end.

A direct application of the general rule yields

$$P(b|j, \neg m) = \alpha \sum_{E} \sum_{A} P(b, j, \neg m, E, A)$$

Inference by Enumeration

Observe that the summands are joint probabilities of all the variables. Hence, we introduce the global semantics of the network:

$$P(b|j, \neg m) = \alpha \sum_E \sum_A P(b)P(E)P(A|b,E)P(j|A)P(\neg m|A)$$

Expanding by enumerating the summands we obtain

$$\begin{aligned} P(b|j,\neg m) = \alpha\,[&P(b)P(e)P(a|b,e)P(j|a)P(\neg m|a) \\ &+P(b)P(e)P(\neg a|b,e)P(j|\neg a)P(\neg m|\neg a) \\ &+P(b)P(\neg e)P(a|b,\neg e)P(j|a)P(\neg m|a) \\ &+P(b)P(\neg e)P(\neg a|b,\neg e)P(j|\neg a)P(\neg m|\neg a)] \end{aligned}$$

For each of the components on the RHS, we read its value from the appropriate CPT, yielding $P(b|j,\neg m) = \alpha 0.00025677$.

Note that the result does not yet amount to a probability value as we haven't solved for α.

Inference by Enumeration (cont.)

To compute $\alpha = \frac{1}{P(j,\neg m)}$ we obtain the marginal probability

$$P(j,\neg m) = \sum_B \sum_E \sum_A P(B,E,A,j,\neg m)$$

An alternative is to realise that $\langle P(b|j,\neg m), P(\neg b|j,\neg m)\rangle$ is a probability distribution and that α is a normalisation constant that ensures that the probability distribution sums to 1.

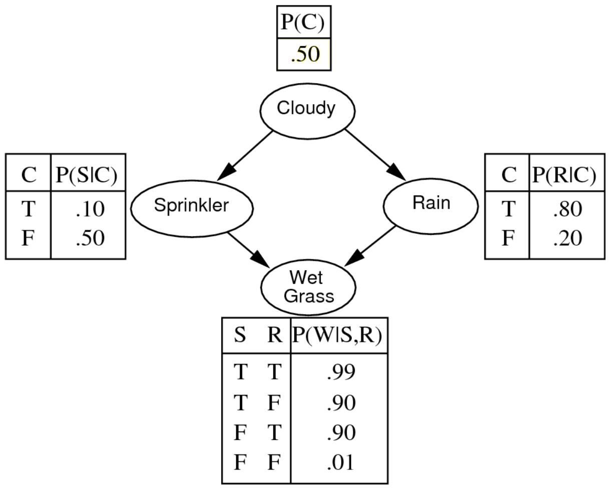
Hence, compute $P(\neg b|j,\neg m) = \alpha 0.0498$, using the global semantics and enumerating the summands as before, yielding

$$\langle P(b|j,\neg m), P(\neg b|j,\neg m)\rangle = \alpha\langle 0.00025677, 0.0498\rangle = \langle 0.0051, 0.9949\rangle$$

where α is solved as $\frac{1}{0.00025677+0.0498}$.
The answer to our query: If John called and Mary didn't most probably there isn't a burglary.

What is the probability of burglary if both John and Mary called?

Wet grass example



P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

You come home one day and see that the grass is wet. What is the probability that it was a cloudy day?

Wet grass example (cont.)

Let the variables be abbreviated as W ($Wetgrass$), S ($Sprinkler$), R ($Rain$) and C ($Cloudy$).

We compute

$$\begin{aligned} P(c|w) &= \alpha \sum_S \sum_R P(c,w,S,R) \\ &= \alpha \sum_S \sum_R P(c)P(S|c)P(R|c)P(w|S,R) \\ &= \alpha\,[P(c)P(s|c)P(r|c)P(w|s,r) \\ &\quad +P(c)P(s|c)P(\neg r|c)P(w|s,\neg r) \\ &\quad +P(c)P(\neg s|c)P(r|c)P(w|\neg s,r) \\ &\quad +P(c)P(\neg s|c)P(\neg r|c)P(w|\neg s,\neg r)] \\ &= \ldots \quad \text{(substitute values from the CPTs)} \end{aligned}$$

The explaining away effect

Back to the Burglar Alarm network. Given that the alarm is ringing, the probability of burglary is

$$P(b|a) = \alpha \sum_J \sum_M \sum_E P(b, a, J, M, E) = 0.3736$$

If we additionally know that there was an earthquake, the probability of burglary becomes

$$P(b|a, e) = \alpha \sum_J \sum_M P(b, a, e, J, M) = 0.0033$$

This makes intuitive sense: If we know there was an earthquake we would immediately attribute the alarm to it. In other words, the earthquake "explains away" the role of burglary.

(Verify the above probability values as exercise) (The probability that there is burglary given that the alarm is ringing is quite low at 0.3736. Why?)

The explaining away effect (cont.)

Using the wet grass example, given that the grass is wet the probability that the sprinkler was on is

$$P(s|w) = 0.43$$

On the other hand, the probability that it was raining given the grass is wet is

$$P(r|w) = 0.708$$

However, if we know that it was indeed raining and the grass is wet, the probability that the sprinkler was on becomes

$$P(s|w, r) = 0.1945$$

i.e. the evidence of rain explains away the cause of the sprinkler.

(Verify the above probability values as exercise)

Complexity of inference by enumeration

In the burglar-alarm example, we evaluate the expression

$$P(b|j, \neg m) \quad = \quad \alpha \sum_E \sum_A P(b)P(E)P(A|b, E)P(j|A)P(\neg m|A)$$

by adding up 4 terms, each obtained by multiplying 5 numbers— In total we need 16 multiplications and 3 additions (excludes the contribution due to term α).

In the worst case, where we have to sum out almost all of the n variables (where we assume they are all Boolean), the complexity of inference by enumeration is $\mathcal{O}(n2^n)$.

This means we will not be able to perform inference by enumeration except for the smallest networks!

Depth-first evaluation

An improvement can be achieved by observing that $P(b)$ is a constant that can be moved outside the summations over E and A, while $P(e)$ can be moved outside the summation over A:

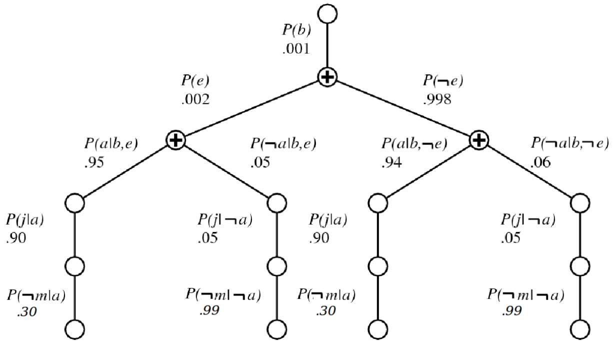
$$P(b|j, \neg m) = \alpha P(b) \sum_E P(E) \sum_A P(A|b, E)P(j|A)P(\neg m|A)$$

Thus we evaluate by looping through the variables in order, multiplying CPT entries as we go.

Now evaluating the expression involves 11 multiplications and 3 additions (again excluding the contribution due to term α).

Depth-first evaluation (cont.)

This method be summarised by the evaluation tree



which shows that we are essentially performing a depth-first evaluation (top to bottom), multiplying and adding terms as we go.

This improves the complexity to $\mathcal{O}(2^n)$— better than $\mathcal{O}(n2^n)$ but still rather grim.

Note also the repeated sub-trees: $P(j|a)P(\neg m|a)$ and $P(j|\neg a)P(\neg m|\neg a)$ are computed twice!

Performance of exact inference methods

More efficient algorithms exist for performing exact inference. For example, the variable elimination method is designed specifically to avoid repeated sub-trees encountered in depth-first evaluation (see Section 14.5 Russell and Norvig).

In singly connected networks (or polytrees) where there is at most 1 (undirected) path between any 2 nodes in the network (e.g. the burglar-alarm network), exact inference algorithms has complexity which is linear in the size of the network.

In the general case where we have multiply connected networks (e.g. the wetgrass network) exact inference can have exponential complexity in the worst case, even if we use variable elimination.

Thus exact inference is impractical for most real-life problems. To perform inference on larger and more complex Bayesian Networks, we will study approximate inference methods in the next lecture.