Informatics 2D: Reasoning and Agents

Alex Lascarides

School of Informatics
The University of Edinburgh

Lecture 23b: The Semantics of Bayesian Networks

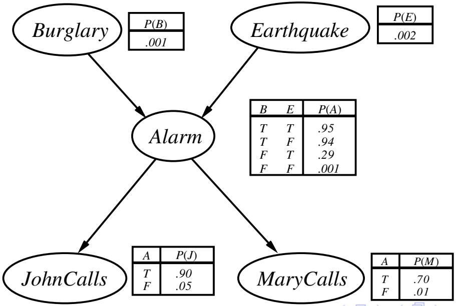
Where are we?

Last time...

- An introduction to Bayesian Networks:
  1. Graph of random variables: arcs showing dependencies (more today)
  2. $P(X|Parents(X))$ for each random variable X
- Today: **Semantics of Bayesian Networks**
  Compact representation of JPDs

Reminder of the example BN

- New burglar alarm has been fitted, fairly reliable but sometimes reacts to earthquakes
- Neighbours John and Mary promise to call when they hear alarm
- John sometimes mistakes phone for alarm, and Mary listens to loud music and sometimes doesn't hear alarm



The semantics of Bayesian Networks

- Two views:
  - BN as representation of JPD (useful for constructing BNs)
  - BN as collection of conditional independence statements (useful for designing inference procedures)
- Every entry $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$ in the JPD can be calculated from a BN (abbreviate by $P(x_1, \ldots, x_n)$)
- $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$
- Example:

$$\begin{aligned} &P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) \\ &\quad = P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e) \\ &\quad = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062 \end{aligned}$$

- As before, this can be used to answer any query

A method for constructing BNs

- Recall product rule for n variables:

$$P(x_1,\ldots,x_n) = P(x_n|x_{n-1},\ldots,x_1)P(x_{n-1},\ldots,x_1)$$

- Repeated application of this yields the so-called **chain rule**:

$$\begin{aligned} P(x_1,\ldots,x_n) &= P(x_n|x_{n-1},\ldots,x_1)P(x_{n-1}|x_{n-2},\ldots,x_1)\cdots P(x_2|x_1)P(x_1) \\ &= \prod_{i=1}^{n} P(x_i|x_{i-1},\ldots,x_1) \end{aligned}$$

- With this we obtain $\mathbf{P}(X_i|X_{i-1},\ldots,X_1) = \mathbf{P}(X_i|Parents(X_i))$ as long as $Parents(X_i) \subseteq \{X_{i-1},\ldots,X_1\}$ (this can be ensured by labelling nodes appropriately)
- For example, it is reasonable to assume that

$$\mathbf{P}(MaryCalls|JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls|Alarm)$$

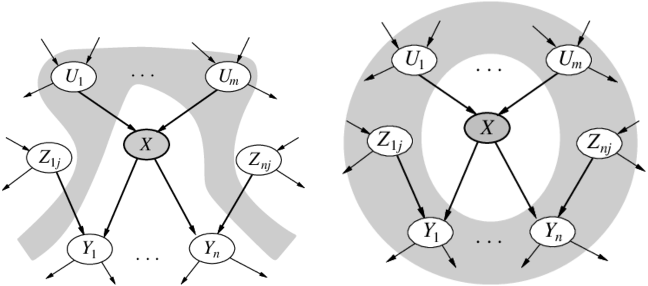
Compactness and node ordering

- BNs examples of **locally structured (sparse)** systems: subcomponents only interact with small number of other components
- E.g. if 30 nodes and every node depends on 5 nodes, BN will have $30 \times 2^5 = 960$ probabilities stored in the CPDs, while JPD would have $2^{30} \approx 1000^3$ entries
- But remember that this is based on designer's independence assumptions!
- Also not trivial to determine good BN structure:

  Add "root causes" first, then variables they influence, and so on, until we reach "leaves" which have no influence on other variables

Conditional independence relations in BNs

- Have provided "numerical" semantics, but can also look at (equivalent) "topological" semantics, namely:
  1. A node is conditionally independent of its **non-descendants**, given its parents
  2. A node is conditionally independent of all other nodes, given its parents, children and children's parents, i.e. its **Markov blanket**



Summary

- BNs are a compact representation of JPDs
- They capture, and so enable us to exploit, conditional independence among the random variables
- Next time: **More on CPTs in BNs**