The exact method for constructing Bayesian networks

Let \( V \) be a finite set of finite propositional variables, \((\Omega, F, P)\) be their joint probability distribution, and \( G = (V, E) \) be a dag.

For each \( v \in V \), let \( c(v) \) be the set of all parents of \( v \) and \( d(v) \) be the set of all descendents of \( v \). Furthermore, for \( v \in V \), let \( a(v) \) be \( V \setminus \{d(v) \cup \{v\}\} \), i.e., the set of propositional variables in \( V \) excluding \( v \) and \( v \)'s descendents.

Suppose for every subset \( W \subseteq a(v) \), \( W \) and \( v \) are conditionally independent given \( c(v) \); that is, if \( P(c(v)) > 0 \), then

\[
P(v \mid c(v)) = 0 \text{ or } P(W \mid c(v)) = 0 \text{ or } P(v \mid W \cup c(v)) = P(v \mid c(v)).
\]

Then, \( C = (V, E, P) \) is called a Bayesian network [Neapolitan, 1990].
The method \[ \text{[Russell & Norvig, Ch. 14]} \]

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

This choice of parents guarantees the global semantics:

\[
P(X_1, \ldots, X_n) = \Pi_{i=1}^n P(X_i|X_1, \ldots, X_{i-1}) \quad \text{(chain rule)}
\]
\[
= \Pi_{i=1}^n P(X_i|\text{Parents}(X_i)) \quad \text{(by construction)}
\]

An example \[ \text{[Russell & Norvig, Ch. 14]} \]

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: \textit{Burglar}, \textit{Earthquake}, \textit{Alarm}, \textit{JohnCalls}, \textit{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
Suppose we choose the ordering $M, J, A, B, E$.

1. $P(JohnCalls) = P(JohnCalls | MaryCalls)$?
   - No

2. $P(Alarm) = P(Alarm | MaryCalls)$?
   - No
   - $P(Alarm | JohnCalls)$
   - No
   - $P(Alarm | JohnCalls, MaryCalls) = P(Alarm | MaryCalls)$?
   - No

3. $P(Burglary | Alarm) = P(Burglary | Alarm, MaryCalls, JohnCalls)$?
   - Yes

4. $P(Earthquake | Alarm) ≠ P(Earthquake | JohnCalls, A, B) = P(Earthquake | A, B)$. 

Choose instead $(B, E, A, M, J)$.

An order in which the edges are directed causally always results in a sparser network.

*Empirical observation*
Result with CPTs.
Bad, b/c the state space of Word is too large.