# Analysis tools for Bayesian networks

Laur Tooming, Meelis Kull
University of Tartu, 25 March 2009

Based on chapter 5 of “Bayesian Networks and Decision Graphs” by F. V. Jensen and T. D. Nielsen

# Outline

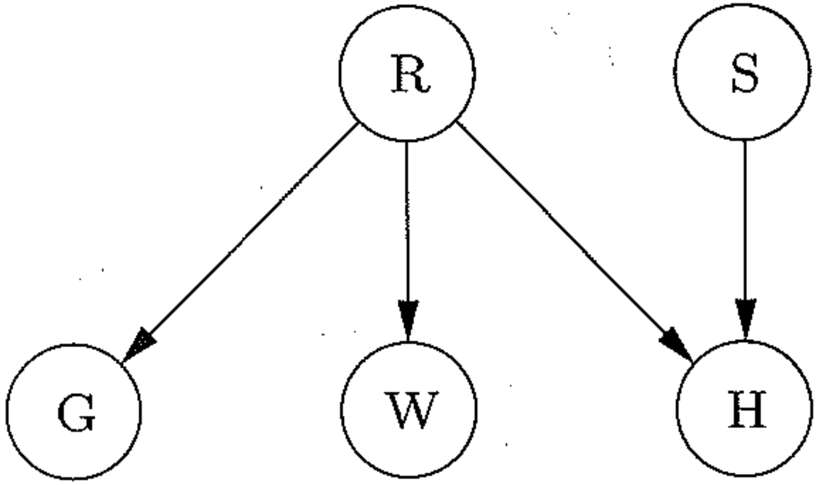
1. Revision
2. IEJ-trees
   - using a part of the evidence after doing computations for the whole evidence
3. Saturated junction trees
   - marginal distributions of several varaibles
4. Configuration of maximal probability
   - What is the likeliest state of the world?
5. Data conflict
   - How to find out if our model doesn’t work
6. SE analysis
   - if we get a result, what part of the evidence contributed most to it?
7. Sensitivity to parameters
   - what happens if we change the parameters of the model?
8. Axioms for propagation in junction trees

# An example

In the morning when Mr Holmes leaves his house, he realizes that his lawn is wet. He wonders whether it has rained during the night or whether he has forgotten to turn off his sprinkler. He looks at the lawn of his neighbors, Dr Watson and Mrs Gibbon. Both lawns are dry, and he concludes that he must have forgotten to turn off his sprinkler.

# An example

- R – rain, {yes,no}
- S – sprinkler, {yes,no}
- H, W, G – Holmes's, Watson's, mrs Gibbon's lawns, {wet (y),dry (n)}




- Evidence: $e_H$=(1,0), $e_G$=(0,1), $e_W$=(0,1)

| | $R = y$ | $R = n$ |
|---|---|---|
| $G = y$ | 0.99 | 0.1 |
| $G = n$ | 0.01 | 0.9 |

$P(G \mid R) = P(W \mid R)$

| | $R = y$ | $R = n$ |
|---|---|---|
| $S = y$ | (1, 0) | (0.9, 0.1) |
| $S = n$ | (0.99, 0.01) | (0, 1) |

$P(H \mid R, S)$

**Table 5.3.** Tables for the wet lawn example. $P(R) = (0.1, 0.9) = P(S)$.

# Revision: junction tree

- Our first goal: calculating the marginal distribution (both prior and posterior) of each variable.
- $P(U)=\prod P(X \mid pa(X))$
- Problem: $P(U)$ is an n-dimensional table, with exponential number of entries
- $P(A) = \Sigma_{U\setminus\{A\}} P(U) = \Sigma_{U\setminus\{A\}} \prod P(X \mid pa(X))$
- In the sum, we can factor out some common factors in the factors
- The junction tree helps us to select the best way of factoring

# Revision: junction tree

- Moral graph: connect each node’s parents
- Junction tree: nodes are sets of variables whose joint probability tables we compute
- Algorithm:
  - Find clique in moral graph that has simplicial nodes
  - Make this clique a node of the junction tree
  - Remove simplicial nodes of clique from moral graph
  - Add nodes of clique that weren’t removed into a separator connected with the new node
  - Repeat until moral graph is empty
  - Connect each separator to node containing its variables
  - Put each potential to node containing its variables

# Revision: propagation

- Choose a root node in the junction tree
- Each separator has 2 mailboxes, for both directions of edges
- Send messages towards root
- Send messages away from root
- Sending message from V to W (separator S):
  - take all potentials at node V
  - take all messages coming in to V, except from S
  - multiply potentials and marginalise out variables V\S
  - put result into S's mailbox exiting V

# Revision: propagation

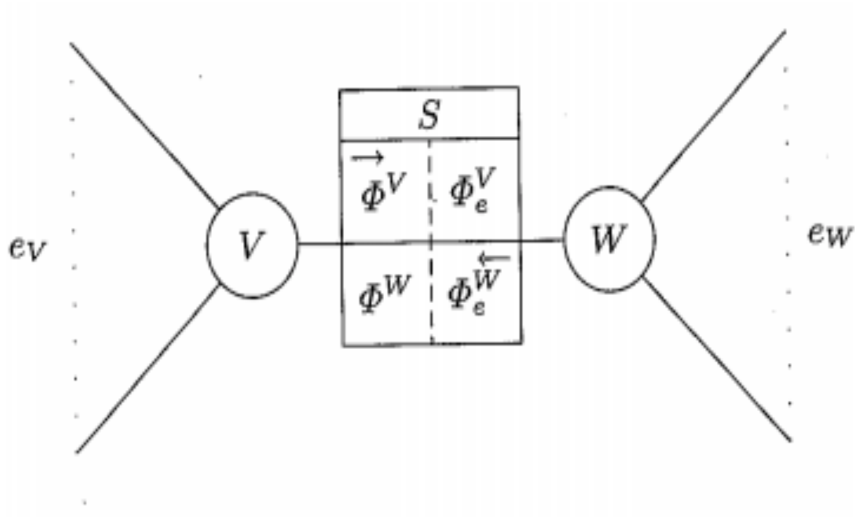
- By multiplying potentials in node V and incoming mailboxes, we get P(V,e)
- P(V|e) can be obtained by normalising
- By multiplying messages at the two mailboxes of a separator S, we get P(S,e)

# IEJ trees

- Purpose: finding probabilities conditioned on *subsets* of the initial evidence set,
- with only 2 propagations (not all subsets have to become available).
- **I**nitial **E**vidence **J**unction-tree – each separator has 4 mailboxes.
- Two propagations:
  – first without any evidence
  – second with all the evidence

# IEJ trees

- From 1st propagation
  - $\Phi^V * \Phi^W = P(S)$ (probability table for all variables at separator S)
- From 2nd propagation
  - $\Phi^V_e * \Phi^W_e = P(S,e)$
- Combined
  - $\Phi^V_e * \Phi^W = P(S,e_V)$ – for evidence entered *left of S*
  - From $P(S,e_V)$ we easily get $P(e_V)$ (how?) and $P(S|e_V)$ (how?)



# Saturated junction trees

- From normal junction trees we can’t find joint marginal distributions like P(A,B|e), if A and B aren’t in a common clique.
- *A-saturated junction tree*: a junction tree where we don’t eliminate the (set of) variable(s) A
- Propagation doesn’t need more time, but needs more space (high-dimensional tables)

# Saturated junction trees

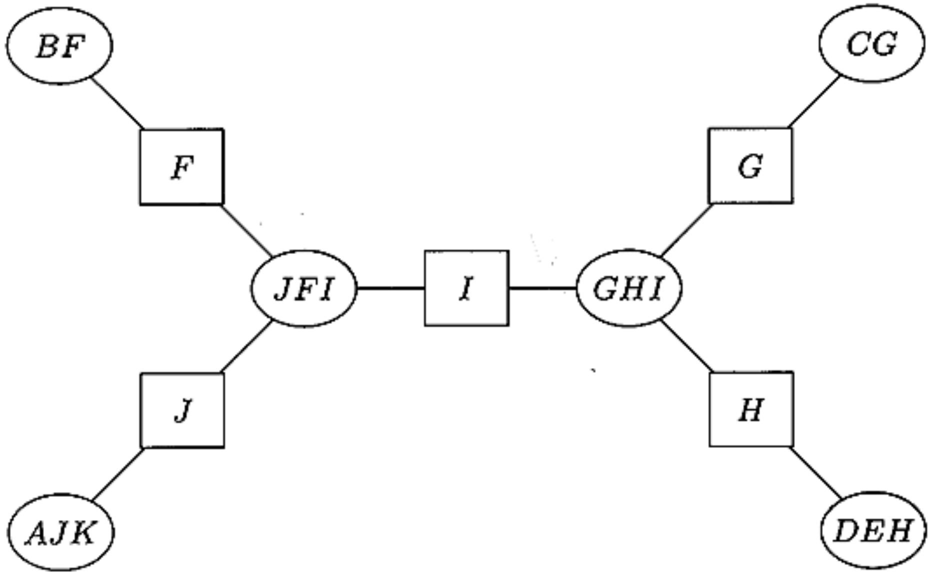



**Fig. 5.2.** A junction tree from which we request $P(A, B, C, D, E)$.

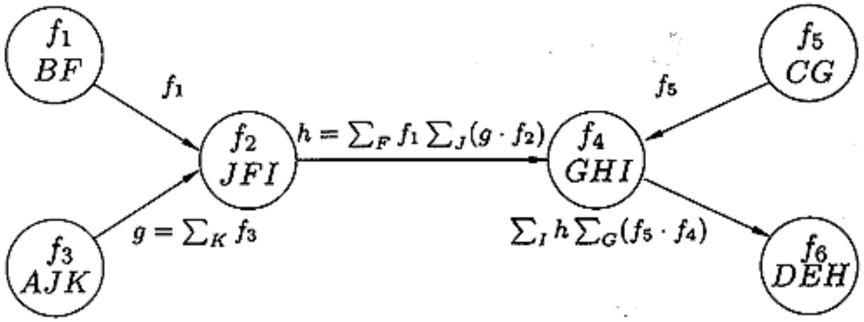



**Fig. 5.3.** The messages passed in performing variable propagation for the calculation of $P(A, B, C, D, E)$. We assume that each clique holds one function (over its domain).

# Configuration of maximal probability

- Which configuration of variables has the greatest probability, given evidence?
- Distributive law for max:

$$\max_Z f(X,Y)g(Y,Z) = f(X,Y)\max_Z g(Y,Z).$$

- max behaves like addition, so the junction tree just works by replacing addition with max

# Configuration of maximal probability

## Theorem 5.1

Suppose we have a Bayesian network and a corresponding junction tree with evidence attached.

Let's do a full round of propagation, using max everywhere instead of summation (*max-propagation*).

Let $U$ denote the set of all variables.

i) For each separator S, $\max_{U \setminus S} P(U,e)$ is the product of potentials in the two mailboxes of S

ii) For each node V, $\max_{U \setminus V} P(U,e)$ is the product of V's potentials and incoming messages to V.

# Configuration of maximal probability

The most probable configuration can be found if it is unique.

- What is the state of variable A?
- Find a separator S or node V containing A.
- Compute $\max_{U\backslash V} P(V,e)$.
- Marginalise on A, getting $\max_{U\backslash\{A\}} P(V,e)$.
- Take the state of A with the highest value.

# Configuration of maximal probability

If there are several maximal configurations:

- for some variables, we get several maximal states
- example: a1 and a2 for A and b1, b2, b3 for B
- unfortunately, not all combinations of those have to be maximal. For example, {a1,b1}, {a1,b2} and {a2,b3} may be maximal while {a1,b3} isn’t.
- To find out, we can enter the states of A as evidence and find the best state for B.

# Configuration of maximal probability

- Suppose we want to find the configuration of maximal probability for a set of variables *Q*, we have evidence for a set of variables *E*, but there are also some other variables.
- Then we have to eliminate some variables by summation and then some by max.
- The result is called *maximum posterior probability*.

$$\max_{\mathcal{Q}} P(\mathcal{Q} \mid e) = \max_{\mathcal{Q}} \sum_{\mathcal{U} \setminus (\mathcal{Q}, \mathcal{E})} P(\mathcal{U} \mid e).$$

- This is difficult to work with, due to constraints on elimination order (first sum, then maximum)

# Data conflicts

- Example: applying a pregnancy test on a man's blood
- Such problems may be diagnosable if we see our evidence is conflicting
- Conflict measure: $\mathrm{conf}(\{e_1, \ldots, e_m\}) = \log_2 \frac{P(e_1) \cdots P(e_m)}{P(e)}$
- If this is positive, there may be a conflict.

# Data conflicts

- A positive conf($e_1,...,e_m$) is *explained* by hypothesis H=h if conf($e_1,...,e_m$,H=h)<=0
- This is equivalent to $$\log_2 \frac{P(h \mid e)}{P(h)} \geq \mathrm{conf}(e)$$

- Example: H=wet, W=G=dry has positive conflict value, but this is explained by S=on
- If there is no such h, we probably have a conflict

# Data conflicts – tracing of conflicts

- Which part of the evidence causes the conflict? Find conf(e') for subsets e' of e.
- Are two subsets of evidence in conflict?

$$\mathrm{conf}(e', e'') = \log_2 \frac{P(e')P(e'')}{P(e)}$$

# Data conflicts – tracing of conflicts

- IEJ trees are used here to deal with subsets of evidence.
- Local conflict at a separator: $\mathrm{conf}(e_V, e_W) = \log_2 \frac{P(e_V)P(e_W)}{P(e)}$
- **Proposition 5.4**

$$\mathrm{conf}(e) = \mathrm{conf}(e_V, e_W) + \mathrm{conf}(e_V) + \mathrm{conf}(e_W).$$