Solutions for Bayesian networks and decision graphs (second edition)

Finn V. Jensen and Thomas D. Nielsen

May 19, 2008

Solution for exercise 7.1

We start off with the complete graph in Figure 1, and perform the independence tests:

- $I(A,B)$, $I(A,C)$, $I(A,D)$, $I(A,E)$, $I(A,F)$, $I(B,C)$, $I(B,D)$, $I(B,E)$, $I(B,F)$, $I(C,D)$, $I(C,E)$, $I(C,F)$, $I(D,E)$, $I(D,F)$, $I(E,F)$.

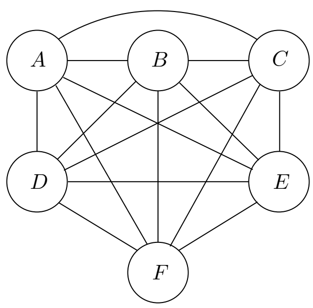

Figure 1: The complete graph for the six variables and the starting point for the PC algorithm.

Out of these tests we find that the oracle answers "yes" to $I(A,C)$, $I(A,E)$, $I(A,F)$, $I(C,D)$ and $I(C,F)$. By removing the corresponding links we get the structure in Figure 2.

Next we condition on one variable and perform independence tests like:

- $I(A,B,D)$, $I(B,C,A)$, $I(B,C,D)$, $I(B,C,E)$, $I(B,C,F)$, $I(B,D,A)$, ...

and we find that the oracle answers "yes" to $I(D,E,F)$.

Using the resulting structure (see Figure 3) we now condition on two variables and ask independence questions like:

- $I(B,A\{D,C\})$, $I(B,A\{C,E\})$, $I(B,A\{C,F\})$, $I(B,A\{D,E\})$, ...

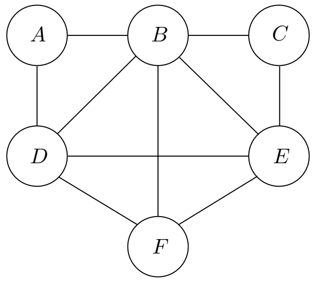

Figure 2: After 1 iteration.

By consulting the oracle we find that $I(A, B, \{D, E\})$, $I(B, C, \{D, E\})$ and $I(B, F, \{D, E\})$ are confirmed, thereby producing the graph in Figure 4.

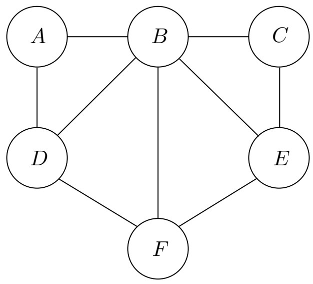

Figure 3: After 2 iterations.

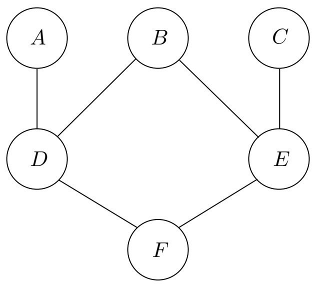

Figure 4: After 3 iterations.

After these three iterations there are no variables with four neighbors, hence the algorithm terminates. By directing the links according to the independences found and using rule 1 we get the structure in Figure 5. Observe that it is not necessary to apply any of the remaining rules, and that, not unexpectedly, the resulting structure is identical to the oracle.

Solution for exercise 7.2

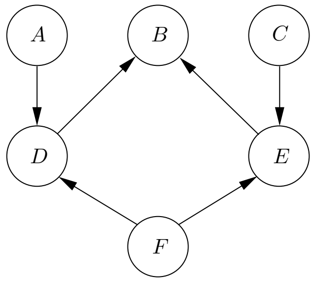

Figure 5: The final model after applying rule 1.

Rule 1: $E \rightarrow I$
Rule 2: $I \rightarrow F$
Rule 3: $C \rightarrow F$
Rule 2: $I \rightarrow H$
Rule 2: $H \rightarrow D$
Rule 2: $D \rightarrow A$
Rule 3: $B \rightarrow A$

Solution for exercise 7.3

The skeleton consists of the following links: $A-E, B-C, B-D, C-E, D-E$.

The result may be any directed graph with the single v-structure $D \rightarrow B \leftarrow C$.

Solution for exercise 7.4

i) The maximal number of independence tests have to be performed when there are no conditional independencies at all. That is, every node has all other nodes as neighbors. In that case we shall for each pair perform an independence test for each subset of remaining nodes. With n nodes there are $\frac{1}{2}n(n-1)$ pairs, and each pair has $n-2$ neighbors. That gives 2^{n-2} conditioning sets. Altogether we get $n(n-1)2^{n-3}$ independence tests.

ii) This is only some input to a discussion. The concern will be to reduce the number of traversals through the database. When an $I(A, B, \mathcal{X})$ is tested we need $P(A, B, \mathcal{X})$ or rather, a contingency table for $\#(A, B, \mathcal{X})$. An efficient algorithm may in a systematic way establish a contingency table as large as possible and exploit it to perform all possibly required tests before a new table is read in. If a test $I(A, B, \mathcal{X})$ requires a contingency table which cannot be held in memory, the algorithm becomes quite different.

Solution for exercise 7.5

Assume A and B to be independent given $\mathcal{X}$. Then

$$P^{\#}(A, B|\mathcal{X}) = P^{\#}(A|\mathcal{X})P^{\#}(B|\mathcal{X}),$$

and clearly $CMI(A, B|\mathcal{X}) = 0$.

To prove in the other direction, we prove that CMI cannot be negative and it can only be 0 if $P^{\#}(A, B|\mathbf{x}) = P^{\#}(A|\mathbf{x})P^{\#}(B|\mathbf{x})$ for every configuration $\mathbf{x}$ of $\mathcal{X}$. Let $P_1(A, B) = P^{\#}(A, B|\mathbf{x})$ and $P_2(A, B) = P^{\#}(A|\mathbf{x})P^{\#}(B|\mathbf{x})$. Then the Kullback-Leibler divergence between P_1 and P_2 is exactly the inner expression of equation 7.2 with $\mathcal{X}$ instantiated to $\mathbf{x}$. As the Kullback-Leibler divergence is never negative, and it can only be zero when the two distributions are identical, we get the result.

Solution for exercise 7.6

For model (a) we have $size_a = |\mathrm{sp}(D)| + |\mathrm{sp}(A)|\,|\mathrm{sp}(D)| + |\mathrm{sp}(A)|\,|\mathrm{sp}(B)|\,|\mathrm{sp}(C)| + |\mathrm{sp}(B)|$ and for model (c) we have $size_c = |\mathrm{sp}(D)|\,|\mathrm{sp}(C)| + |\mathrm{sp}(B)| + |\mathrm{sp}(A)|\,|\mathrm{sp}(B)|\,|\mathrm{sp}(C)|\,|\mathrm{sp}(D)| + |\mathrm{sp}(B)|$. By considering the difference $size_c - size_a$ we get $[|\mathrm{sp}(D)|\,|\mathrm{sp}(C)| - |\mathrm{sp}(D)|] + [|\mathrm{sp}(A)|\,|\mathrm{sp}(B)|\,|\mathrm{sp}(C)|\,|\mathrm{sp}(D)| - |\mathrm{sp}(D)|\,|\mathrm{sp}(A)| + |\mathrm{sp}(C)|\,|\mathrm{sp}(B)| - |\mathrm{sp}(A)|\,|\mathrm{sp}(B)|\,|\mathrm{sp}(C)|]$. The first term is strictly positive, and the second term can be rewritten as $|\mathrm{sp}(A)|\,|\mathrm{sp}(D)|\,[|\mathrm{sp}(B)|\,|\mathrm{sp}(C)| - 1 + |\mathrm{sp}(C)|\,|\mathrm{sp}(B)|\,/(|\mathrm{sp}(A)|\,|\mathrm{sp}(D)|) - |\mathrm{sp}(B)|\,|\mathrm{sp}(C)|\,/\,|\mathrm{sp}(D)|]$. The term inside parenthesis can again be rewritten as $|\mathrm{sp}(B)|\,|\mathrm{sp}(C)|\,[1 - 1/(|\mathrm{sp}(B)|\,|\mathrm{sp}(C)|) + 1/(|\mathrm{sp}(A)|\,|\mathrm{sp}(D)|) - 1/\,|\mathrm{sp}(D)|]$ and by assuming that all variables have at least two states the proof follows.

Solution for exercise 7.7

$$\mathrm{size} = 2 + 2 + 3 \cdot 2 + 3 \cdot 3 \cdot 2 + 3 \cdot 3 \cdot 2 = 46$$

Solution for exercise 7.8

The size of model S_1 is $\mathrm{size}_{S_1} = 1 + 2 = 3$. Furthermore, by calculating frequencies in the database we get

$$P^{\#}(A = 1) = \frac{14}{32};$$
$$P^{\#}(B = 1 \,|\, A = 1) = \frac{11}{14};$$
$$P^{\#}(B = 1 \,|\, A = 2) = \frac{18}{18},$$

form which the log-likelihood of the model can be calculated:

$$\begin{aligned}\log P(\mathcal{D} \mid \hat{\boldsymbol{\theta}}_{S_1}, S_1) =& \left[11 \cdot \log \frac{11}{14} + 14 \cdot \log \frac{14}{32}\right] + \left[3 \cdot \log \frac{3}{14} + 14 \cdot \log \frac{14}{32}\right] \\ &+ \left[18 \cdot \log \frac{18}{18} + 18 \cdot \log \frac{18}{32}\right] \\ =& - 58.8297.\end{aligned}$$

The BIC score is therefore $\mathrm{BIC}(S_1 \mid \mathcal{D}) = -58.8297 - 3/2 \cdot \log(32) = -66.3297$.

The size of model S_2 is $\mathrm{size}_{S_2} = 1 + 2 + 2 = 5$, and in addition to the frequency counts above, we also need

$$\begin{aligned}P^{\#}(C = 1) &= \frac{23}{32}; \\ P^{\#}(A = 1 \mid C = 1) &= \frac{11}{23}; \\ P^{\#}(A = 1 \mid C = 2) &= \frac{3}{9}.\end{aligned}$$

The log-likelihood is therefore given by -106.9986, and the BIC score is -119.4986.

Solution for exercise 7.9

First we estimate the maximum likelihood parameters of the model using the EM algorithm (for the associated net-file the network were initialized with even distributions). Next we calculate the probabilities of the five cases in the database:

$$\begin{aligned}P(1) &= 0.3332; \\ P(2) &= 0.3330; \\ P(3) &= 0.5000; \\ P(4) &= 0.1667; \\ P(5) &= 0.5000,\end{aligned}$$

which give the BIC score $\mathrm{BIC}(M \mid \mathcal{D}) = -7.7566 - \frac{5}{2} \log_2(5) = -13.5614$.

Solution for exercise 7.10

Solution for exercise 7.11 The probability of the data can be written as

$$\begin{aligned}
P(\mathcal{D} \,|\, \hat{\boldsymbol{\theta}}) &= \prod_{l=1}^{N} P(\mathbf{d}_l \,|\, \hat{\boldsymbol{\theta}}) \\
&= \prod_{l=1}^{N} \prod_{i=1}^{n} P(X_i = \mathbf{d}_l^{\downarrow X_i} \mid \mathrm{pa}(X_i) = \mathbf{d}_l^{\downarrow \mathrm{pa}(X_i)}, \hat{\boldsymbol{\theta}}) \\
&= \prod_{l=1}^{N} \prod_{i=1}^{n} \frac{N(X_i = \mathbf{d}_l^{\downarrow X_i}, \mathrm{pa}(X_i) = \mathbf{d}_l^{\downarrow \mathrm{pa}(X_i)})}{N(\mathrm{pa}(X_i) = \mathbf{d}_l^{\downarrow \mathrm{pa}(X_i)})} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \left[\frac{N(X_i = k, \mathrm{pa}(X_i) = j)}{N(\mathrm{pa}(X_i) = j)} \right]^{N(X_i = k, \mathrm{pa}(X_i) = j)},
\end{aligned}$$

and from the definition of size (Proposition 7.1) the proof follows immediately.

Solution for exercise 7.12 Keep a list of the last h moves, and only consider the legal moves that are not on the list.

Solution for exercise 7.13

1. Let $S = S_0$ be the initial structure.
2. Set T to some starting temperature.
3. **repeat**
   (a) Let A be a legal arc operation and let $\Delta(A)$ be the score difference.
   (b) **if** $\Delta(A) > 0$ **then**
       Set $S := op(S, A)$
   (c) **else** set $S := op(S, A)$ with probability $\exp(\Delta(A)/T)$
   (d) $T := T - 1$;
4. **until** $T = 0$.
5. **return** S.

Solution for exercise 7.14

The following three networks are equivalent to the one in Figure 7.27: $\{(A, B), (C, A), (C, D), (B, D)\}$, $\{(A, B), (A, C), (C, D), (B, D)\}$, and $\{(B, A), (A, C), (C, D), (B, D)\}$.

Solution for exercise 7.15

The mutual information between the three pairs of variables are $MI(A, B) = 0.1209$, $MI(A, C) = 0.0127$, and $MI(B, C) = 0.1906$, hence we get a maximal spanning tree consisting of the edges $\{A, B\}$ and $\{B, C\}$. A directed tree can be found be selecting any of the nodes, and directing the edges away from that node.

Solution for exercise 7.16

For the network model M_3 in Figure 7.28 we have the following statistics $N_{111} = 6, N_{112} = 0, N_{121} = 2N_{122} = 2, N_{21} = 6, N_{22} = 4$, hence we get

$$P(\mathcal{D} \,|\, M_3) = \frac{6!1!2!2!6!4!}{7!5!11!} = 2.0614 \cdot 10^{-6}.$$

1. $P(\mathcal{D}) = P(\mathcal{D}, M_1) + P(\mathcal{D}, M_2) + P(\mathcal{D}, M_3) = 1.8688 \cdot 10^{-6}$.

2.
$$P(M_1 \,|\, \mathcal{D}) = \frac{2.67 \cdot 10^{-6}}{3 \cdot 1.8688 \cdot 10^{-6}} = 0.4762$$
$$P(M_2 \,|\, \mathcal{D}) = \frac{8.75 \cdot 10^{-7}}{3 \cdot 1.8688 \cdot 10^{-6}} = 0.1561$$
$$P(M_3 \,|\, \mathcal{D}) = \frac{2.0614 \cdot 10^{-6}}{3 \cdot 1.8688 \cdot 10^{-6}} = 0.3677$$

3. By assuming that the remaining probability mass is distributed evenly among the other two models, then the prior on the empty structure should be larger than 0.6041 (ensuring that $P(\mathcal{D}, \text{empty structure}) > P(\mathcal{D}, M_i)$, for $i = 1, 3$).

Solution for exercise 7.17

By setting $\kappa = 0.5$ (see page 256) we get the prior distribution $P(X_1 \rightarrow X_2) \propto 0.5^2$, $P(X_1 \leftarrow X_2) \propto 0.5^0$, and $P(X_1\, X_2) \propto 0.5$. Using this prior we have $P(\mathcal{D}, X_1 \rightarrow X_2) = 3.81 \cdot 10^{-7}$, $P(\mathcal{D}, X_1\, X_2) = 2.5 \cdot 10^{-7}$, and $P(\mathcal{D}, X_1 \leftarrow X_2) = 1.1 \cdot 10^{-6}$, and the greedy search therefore returns the model $X_1 \leftarrow X_2$.

Solution for exercise 7.18

See Solution 7.17.