Assignments of the subject "Probabilistic Methods"

Topic 2. Inference in Bayesian networks

Exercise 2.1.

Let be the Bayesian network of 6 binary variables given by the links $\{A \longrightarrow B, B \longrightarrow C, B \longrightarrow D, C \longrightarrow F, D \longrightarrow F, D \longrightarrow G\}$ and the following probabilities:

$$P(+a) = 0'1,$$
$$P(+b|+a) = 0'8,\ P(+b|\neg a) = 0'25,$$
$$P(+c|+b) = 0'7,\ P(+c|\neg b) = 0'35,$$
$$P(+d|+b) = 0'6,\ P(+d|\neg b) = 0'1,$$
$$P(+f|+c,+d) = 0'8,\ P(+f|+c,\neg d) = 0'6,\ P(+f|\neg c,+d) = 0'5,\ P(+f|\neg c,\neg d) = 0,\ P(+g|+d) = 0'4,\ P(+g|\neg d) = 0'1.$$

1. Calculate the probability $P(b|\neg a, +f, \neg g)$ with the variable elimination method, indicating the **numerical values** of the potentials that are calculated in each step.

2. Calculate the same probability using the OpenMarkov program. Capture the screen for this network in inference mode, with the findings entered, and include it in your answers.

Solution 2.1.

1)

To calculate P(b|¬a,+f,¬g) using variable elimination:

1. Initial Factors

Six factors encode the BN's conditional probabilities:

- Factor 1: P(C|B)
- Factor 2: P(G|D)
- Factor 3: P(D|B)
- Factor 4: P(F|C,D)
- Factor 5: P(B|A)
- Factor 6: P(A)

2. Evidence Handling

- Fix A=False, F=True, *G*=False.
- Remove rows inconsistent with evidence from relevant factors.

3. Eliminate C

Multiply factors containing C (Factors 1 and 4) and marginalize C:

P(B, D, F) = ΣC [P(C | B) * P(F | C, D)]

Product of factors containing C

P({'B': False, 'C': True, 'D': False, 'F': True}) = 0.2100

P({'B': True, 'C': True, 'D': False, 'F': True}) = 0.4200

P({'B': False, 'C': False, 'D': True, 'F': True}) = 0.3250

P({'B': True, 'C': False, 'D': True, 'F': True}) = 0.1500

P({'B': False, 'C': True, 'D': True, 'F': True}) = 0.2800

P({'B': True, 'C': True, 'D': True, 'F': True}) = 0.5600

P({'B': False, 'C': False, 'D': False, 'F': False}) = 0.6500

P({'B': True, 'C': False, 'D': False, 'F': False}) = 0.3000

P({'B': False, 'C': True, 'D': False, 'F': False}) = 0.1400

P({'B': True, 'C': True, 'D': False, 'F': False}) = 0.2800

P({'B': False, 'C': False, 'D': True, 'F': False}) = 0.3250

P({'B': True, 'C': False, 'D': True, 'F': False}) = 0.1500

P({'B': False, 'C': True, 'D': True, 'F': False}) = 0.0700

P({'B': True, 'C': True, 'D': True, 'F': False}) = 0.1400

After marginalizing C

P({'B': False, 'D': False, 'F': True}) = 0.2100

P({'B': True, 'D': False, 'F': True}) = 0.4200

P({'B': False, 'D': True, 'F': True}) = 0.6050

P({'B': True, 'D': True, 'F': True}) = 0.7100

P({'B': False, 'D': False, 'F': False}) = 0.7900

P({'B': True, 'D': False, 'F': False}) = 0.5800

P({'B': False, 'D': True, 'F': False}) = 0.3950

P({'B': True, 'D': True, 'F': False}) = 0.2900

<u>4. Eliminate D</u>

Multiply updated factors with D (Factors 2 and 3, marginalized C) and marginalize D:

P(B, F, G) = ΣD [P(D | B) * P(G | D) * prior results]

Product of factors containing D

P({'B': False, 'D': True, 'F': True, 'G': True}) = 0.0242

P({'B': False, 'D': True, 'F': False, 'G': True}) = 0.0158

P({'B': False, 'D': True, 'F': True, 'G': False}) = 0.0363

P({'B': False, 'D': True, 'F': False, 'G': False}) = 0.0237

P({'B': True, 'D': True, 'F': True, 'G': True}) = 0.1704

P({'B': True, 'D': True, 'F': False, 'G': True}) = 0.0696

P({'B': True, 'D': True, 'F': True, 'G': False}) = 0.2556

P({'B': True, 'D': True, 'F': False, 'G': False}) = 0.1044

P({'B': False, 'D': False, 'F': True, 'G': True}) = 0.0189

P({'B': False, 'D': False, 'F': False, 'G': True}) = 0.0711

P({'B': False, 'D': False, 'F': True, 'G': False}) = 0.1701

P({'B': False, 'D': False, 'F': False, 'G': False}) = 0.6399

P({'B': True, 'D': False, 'F': True, 'G': True}) = 0.0168

P({'B': True, 'D': False, 'F': False, 'G': True}) = 0.0232

P({'B': True, 'D': False, 'F': True, 'G': False}) = 0.1512

P({'B': True, 'D': False, 'F': False, 'G': False}) = 0.2088

After marginalizing D

P({'B': False, 'F': True, 'G': True}) = 0.0431

P({'B': False, 'F': False, 'G': True}) = 0.0869

P({'B': False, 'F': True, 'G': False}) = 0.2064

P({'B': False, 'F': False, 'G': False}) = 0.6636

P({'B': True, 'F': True, 'G': True}) = 0.1872

P({'B': True, 'F': False, 'G': True}) = 0.0928

P({'B': True, 'F': True, 'G': False}) = 0.4068

P({'B': True, 'F': False, 'G': False}) = 0.3132

5. Combine Remaining Factors

The formula for the query is:
P(query | evidence) = P(query, evidence) / P(evidence)

Calculating P(query, evidence):
We need to multiply the remaining factors and sum over all assignments consistent with evidence

Final joint factor:

P({'A': True, 'B': True, 'F': True, 'G': True}) = 0.0150

P({'A': True, 'B': True, 'F': False, 'G': True}) = 0.0074

P({'A': True, 'B': True, 'F': True, 'G': False}) = 0.0325

P({'A': True, 'B': True, 'F': False, 'G': False}) = 0.0251

P({'A': True, 'B': False, 'F': True, 'G': True}) = 0.0009

P({'A': True, 'B': False, 'F': False, 'G': True}) = 0.0017

P({'A': True, 'B': False, 'F': True, 'G': False}) = 0.0041

P({'A': True, 'B': False, 'F': False, 'G': False}) = 0.0133

P({'A': False, 'B': True, 'F': True, 'G': True}) = 0.0421

P({'A': False, 'B': True, 'F': False, 'G': True}) = 0.0209

P({'A': False, 'B': True, 'F': True, 'G': False}) = 0.0915

P({'A': False, 'B': True, 'F': False, 'G': False}) = 0.0705

P({'A': False, 'B': False, 'F': True, 'G': True}) = 0.0291

P({'A': False, 'B': False, 'F': False, 'G': True}) = 0.0587

P({'A': False, 'B': False, 'F': True, 'G': False}) = 0.1393

P({'A': False, 'B': False, 'F': False, 'G': False}) = 0.4479

Result:

Adding P({'A': False, 'B': True, 'F': True, 'G': False}) = 0.0915

Adding P({'A': False, 'B': False, 'F': True, 'G': False}) = 0.1393

P(query,evidence) = 0.230850

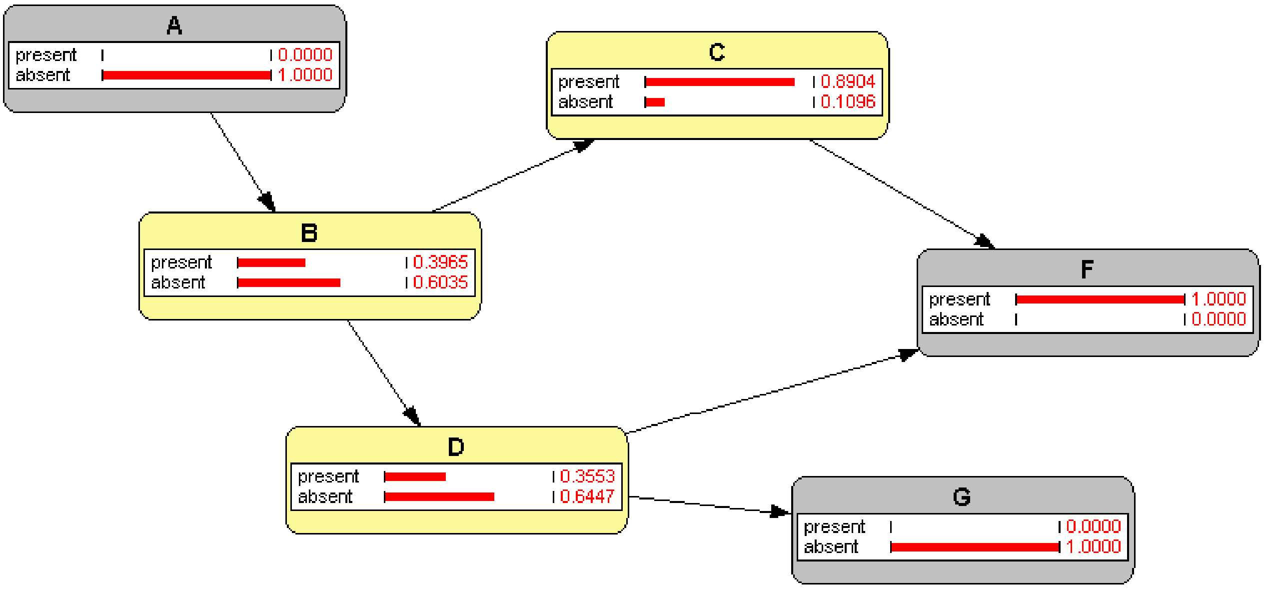
P(evidence) = 0.230850

P(query|evidence) = P(query,evidence) / P(evidence)

P(query|evidence) = 0.091530 / 0.230850 = 0.396491

P(b|¬a, +f, ¬g) = 0.396491

2)



Exercise 2.2.

Indicate in detail what steps each of the following methods would follow to calculate the above probability. You don't need to do numerical calculations.

1. Logical sampling (called the "accept-reject method" in [1]).
2. Weighting by likelihood (called the "weighing likelihood function method" in [1]).

Solution 2.2.

1. Logical Sampling (Accept–Reject)

Logical sampling (also known as the accept–reject method) works by drawing complete samples from the joint and then discarding those that don't match the evidence.

1.1 Choose a topological ordering of the variables

e.g.: A, B, C, D, F, G

1.2 Repeat for NN samples

For i=1i=1 to NN:
a. Sample ai ~ P(A)

b. Sample bi ~ P(B | ai)

c. Sample ci ~ P(C | bi)

d. Sample di ~ P(D | bi)

e. Sample fi ~ P(F | ci, di)

f. Sample gi ~ P(G | di)

1.3 Reject non–evidence samples
If the sampled (ai, bi, ci, di, fi, gi) does not match the evidence
(ai = –, fi = +, gi = –), discard ("reject") it.

1.4 Count accepted samples
Among the remaining ("accepted") samples, count how many have bi = +;
let that count be M^+, and let the total number of accepted samples be M.

1.5 Estimate

$P(B = + \mid E) \approx M^+ / M$

• Note: if P(F = +, G = –) is small, most samples get rejected and efficiency is low.

2. Likelihood Weighting

Likelihood weighting avoids rejection by always respecting the evidence but attaching a weight to each sample that reflects how likely the evidence is under that sample's ancestry.

2.1 Topological ordering
Same as before: A, B, C, D, F, G
2.2 For each weighted sample i = 1 to N:
– Initialize weight wi ← 1
– Initialize an empty assignment for all variables
2.3 For each variable X in order:
– If X is **not** evidence (i.e. X ∉ {A, F, G}):
• Sample xi ~ P(X | Parents(X))
• Leave wi unchanged
– If X **is** evidence (X ∈ {A, F, G}):
• Set xi to the observed value (ai = –, fi = +, gi = –)
• Update weight:
wi = wi * P(X = xi | Parents(X))

2.4 After assigning all variables for sample i, you have bi and weight wi.

– If bi = +, add wi to W^+

– In any case, add wi to the total weight W

2.5 Estimate

$P(B = + \mid E) \approx W^+ / W$

• Since no samples are discarded, likelihood weighting is more efficient when evidence is unlikely.

Exercise 2.3.

Let be the Bayesian network of 12 variables given by the links $\{A \rightarrow D, A \rightarrow F, B \rightarrow F, B \rightarrow G, C \rightarrow G, C \rightarrow H, D \rightarrow I, F \rightarrow I, F \rightarrow J, G \rightarrow J, G \rightarrow K, H \rightarrow K, I \rightarrow L, J \rightarrow L, J \rightarrow M, K \rightarrow M\}$. Indicate in detail the steps required to calculate the a posteriori probability of A given a certain value of L (e.g., +l) using the following methods. [We are supposed to know the conditional probability tables that define the network: $P(a)$, $P(d|a)$, etc., but in the statement we do not give their numerical values because we only want it to explain how potentials are calculated.] If desired, you can delete the sink nodes in each method before you start inference.

1. Elimination of variables.
2. Grouping (also indicate how the tree is constructed).
3. Arc inversion (indicate how the new probability tables are calculated).

Solution 2.3.

Calculating P(A|L=True)

1. **Variable Elimination.**

Variable elimination is a method to compute exact inference in Bayesian networks by eliminating variables one by one. Here's how to calculate P(A|L=+l) for the given network:

Step 1: Incorporate Evidence

First, incorporate the evidence L=+l by reducing the factor P(L|I,J) to a factor ψ(I,J) that only includes values consistent with L=+l.

Step 2: Choose an Elimination Order

Select an elimination order for all variables except A. A good elimination order would be: M, D, H, K, I, C, B, G, F, J.

Step 3: Eliminate Variables One by One

Eliminate M:

- Identify factor containing M: P(M|J,K)
- Since M is a sink node, we can simply ignore this factor
- Result: No new factor created

Eliminate D:

- Identify factors containing D: P(D|A) and P(I|D,F)
- Multiply these factors: P(D|A) × P(I|D,F) = ψ(D,A,I,F)
- Marginalize out D by summing over all its values: ψ(A,I,F)

Eliminate H:

- Identify factors containing H: P(H|C) and P(K|G,H)
- Multiply these factors: P(H|C) × P(K|G,H) = ψ(H,C,K,G)
- Marginalize out H: ψ(C,K,G)

Eliminate K:

- Identify factors containing K: $\psi(C,K,G)$ and any other K factors
- Multiply these factors to get $\psi(J,K,C,G)$
- Marginalize out K: $\psi(J,C,G)$

Eliminate I:

- Identify factors containing I: $\psi(A,I,F)$ and $\psi(I,J)$
- Multiply these factors: $\psi(A,I,F) \times \psi(I,J) = \psi(I,J,A,F)$
- Marginalize out I: $\psi(J,A,F)$

Eliminate C:

- Identify factors containing C: $P(C)$, $P(G|B,C)$, and $\psi(J,C,G)$
- Multiply these factors: $\psi(C,G,B,J)$
- Marginalize out C: $\psi(G,B,J)$

Eliminate B:

- Identify factors containing B: $P(B)$, $P(F|A,B)$, and $\psi(G,B,J)$
- Multiply these factors: $\psi(B,F,A,G,J)$
- Marginalize out B: $\psi(F,A,G,J)$

Eliminate G:

- Multiply relevant factors: $\psi(J,F,G,A)$
- Marginalize out G: $\psi(J,F,A)$

Eliminate F:

- Multiply relevant factors: $\psi(J,A,F)$
- Marginalize out F: $\psi(J,A)$

Eliminate J:

- Marginalize out J from $\psi(J,A)$: $\psi(A)$

Step 4: Normalize

The final factor $\psi(A)$ is proportional to $P(A|L=+l)$. Normalize by dividing by the sum over all values of A:

$P(A|L=+l) = \psi(A) / \Sigma_a \psi(A)$

This completes the variable elimination process to calculate the posterior probability $P(A|L=+l)$.

2. **Grouping**

The grouping method (also known as junction tree or clustering) is a technique for exact inference in Bayesian networks.

Step 1: Create Initial Factors

First, create factors for each conditional probability in the Bayesian network:

- Prior probabilities: P(A), P(B), P(C)
- Conditional probabilities: P(D|A), P(F|A,B), P(G|B,C), P(H|C), P(I|D,F), P(J|F,G), P(K|G,H), P(L|I,J), P(M|J,K)

Step 2: Incorporate Evidence

Reduce the factor P(L|I,J) to account for the evidence L=+l, creating a factor ψ(I,J).

Step 3: Construct the Junction Tree

First, we need to moralize and triangulate the directed graph to create a junction tree:

1. Connect parents of each node (moralization)
2. Add edges to eliminate cycles of length 4 or more (triangulation)
3. Identify maximal cliques as clusters
4. Connect the clusters to form a tree that satisfies the running intersection property

Step 4: Group Factors

Group the factors in the following sequence to create larger factors:

1. Group P(D|A) with P(A):
   - Multiply these factors to create a joint factor ψ(D,A)
2. Group P(H|C) with P(C):
   - Multiply to create ψ(H,C)
3. Group P(G|B,C) with P(B):
   - Multiply to create ψ(G,B,C)
4. Group ψ(I,J) with P(M|J,K):
   - Multiply to create ψ(I,J,M,K)

5. Group $P(K|G,H)$ with $\psi(H,C)$:
    - Multiply to create $\psi(K,G,H,C)$
6. Group $P(F|A,B)$ with $\psi(D,A)$:
    - Multiply to create $\psi(F,A,B,D)$
7. Group $P(J|F,G)$ with $P(I|D,F)$:
    - Multiply to create $\psi(J,F,G,I,D)$
8. Group $\psi(G,B,C)$ with $\psi(K,G,H,C)$:
    - Multiply to create $\psi(G,B,C,K,H)$
9. Group $\psi(I,J,M,K)$ with $\psi(J,F,G,I,D)$:
    - Multiply to create $\psi(I,J,M,K,F,G,D)$
10. Group $\psi(F,A,B,D)$ with $\psi(G,B,C,K,H)$:
    - Multiply to create $\psi(F,A,B,D,G,C,K,H)$
11. Group $\psi(I,J,M,K,F,G,D)$ with $\psi(F,A,B,D,G,C,K,H)$:
    - Multiply to create a comprehensive factor $\psi(I,J,M,K,F,G,D,A,B,C,H)$

Step 5: Message Passing

Perform message passing (belief propagation) between the clusters in the junction tree:

1. Choose a root cluster
2. Pass messages from leaf clusters to the root (collect)
3. Pass messages from root to leaves (distribute)

Step 6: Compute the Posterior

1. From the final grouped factor, marginalize out all variables except A
2. Normalize the resulting factor to obtain $P(A|L=+l)$:
   $P(A|L=+l) = \psi(A) / \Sigma_a \psi(A)$

This grouping approach efficiently exploits the conditional independence relationships in the network to compute the desired posterior probability.

3. Arc inversion

Arc inversion is a method that applies Bayes' rule to reverse directed arcs in a Bayesian network to perform inference. Here's how to calculate P(A|L=+l) using this technique

Step 1: Incorporate Evidence

Reduce the factor P(L|I,J) to account for the evidence L=+l, creating a factor ψ(I,J).

Step 2: Identify Path for Inversion

Identify a directed path from query variable A to evidence variable L. In this network, the path is:
$A \to D \to I \to L$

We'll invert each arc along this path.

Step 3: Invert Arc A→D

Apply Bayes' rule to invert the direction of A→D:

- Original factor: P(D|A)
- New factor: P(A|D) = P(D|A)P(A)/P(D)
  where $P(D) = \sum_a P(D|A)P(A)$

This inversion affects other factors involving A, particularly P(F|A,B). After inversion, we have a combined factor with scope [D,A,F,B,I].

Step 4: Invert Arc D→I

Next, invert D→I:

- Original factor: P(I|D,F)
- New factor: P(D|I,F) = P(I|D,F)P(D)/P(I|F)
  where $P(I|F) = \sum^d P(I|D,F)P(D)$

After inversion, we have a combined factor with scope [I,J,D,A,F,B].

Step 5: Invert Arc I→L

Finally, invert I→L (with evidence L=+l):

- Original factor: P(L|I,J)
- New factor: P(I|L=+l,J) = P(L=+l|I,J)P(I|J)/P(L=+l|J)
  where $P(L=+l|J) = \sum_i P(L=+l|I,J)P(I|J)$

After inverting I→L, we have effectively incorporated the evidence L=+l into our network.

Step 6: Compute Joint Distribution

Multiply all the inverted factors with the remaining original factors to obtain a joint distribution with scope [G,B,C,H,K,M,J,F,I,D,A] conditioned on L=+l.

The multiplication combines:

- The inverted factors we created in steps 3-5
- All original factors for variables not involved in the inversion
-

Step 7: Marginalize and Normalize

1. Marginalize out all variables except A by summing over all their possible values
2. Normalize the resulting factor to ensure the probabilities sum to 1

The final result gives P(A|L=+l).

Note: Arc inversion maintains the exact inference capability while changing the structure of the network to directly model the posterior distribution given the observed evidence.

References

[1] E. Castillo, J. M. Gutiérrez and A. S. Hadi. *Expert Systems and Models of Probabilistic Networks*. Academy of Engineering, Madrid, 1997.