Bayesian Networks

Exact Inference by Variable Elimination

Emma Rollon and Javier Larrosa

Q1-2015-2016

Recall the most usual queries

Let $\mathcal{X} = E \cup Q \cup Z$, where E are the evidence variables, e are the observed values, Q are the variables of interest, Z are the rest of the variables.

- **Probability of Evidence ($Q = \emptyset$):**

$$Pr(e) = \sum_Z Pr(e, Z)$$

- **Prior Marginals** ($E = \emptyset$)**:**

$$Pr(Q) = \sum_Z Pr(Q, Z)$$

- **Posterior Marginals ($E \neq \emptyset$):**

$$Pr(Q \mid e) = \sum_Z Pr(Q, Z \mid e)$$

- **Most Probable Explanation (MPE) ($E \neq \emptyset$, $Z = \emptyset$):**

$$q^* = \arg\max_Q Pr(Q \mid e)$$

- **Maximum a Posteriori Probability (MAP):**

$$q^* = \arg\max_Q Pr(Q \mid e) = \arg\max_Q \sum_Z Pr(Q, Z \mid e)$$

In this session

We are going to see a generic algorithm for inference (inference = query answer)

It is called Variable Elimination because it eliminates one by one those variables which are irrelevant for the query.

- It relies on some basic operations on a class of functions known as factors.
- It uses an algorithmic technique called dynamic programming

We will illustrate it for the computation of *prior* and *posteriors marginals*,

$$P(\mathbf{Y}) = \sum_{Z} P(\mathbf{Y}, Z)$$

$$P(\mathbf{Y}|e) = \sum_{Z} P(\mathbf{Y}, Z|e)$$

- It can be easily extended to other queries.

Factors

A factor f is a function over a set of variables $\{X_1, X_2, \ldots, X_k\}$, called *scope*, mapping each instantiation of these variables to a real number $\mathcal{R}$.

X	Y	Z	$f(X, Y, Z)$
0	0	0	5
0	0	1	3
0	0	2	1
0	1	0	5
0	1	1	2
0	1	2	7
1	0	0	1
1	0	1	4
1	0	2	6
1	1	0	2
1	1	1	4
1	1	2	3

Observation: Condition Probability Tables (CPTs) are factors.

We define three operations on factors:

- Product: $f(X, Y, Z) \times g(Q, Y, R)$
- Conditioning: $f(X, y, Z)$
- Marginalization: $\sum_{y \in Y} f(X, y, Z)$

Factors: Product

X	Y	$f(X,Y)$
0	0	1
0	1	3
1	0	2
1	1	1

Y	Z	$g(Y,Z)$
0	0	4
0	1	3
1	0	1
1	1	2

X	Y	Z	$f(X,Y) \times g(Y,Z)$
0	0	0	$1 \cdot 4$
0	0	1	$1 \cdot 3$
0	1	0	$3 \cdot 1$
0	1	1	$3 \cdot 2$
1	0	0	$2 \cdot 4$
1	0	1	$2 \cdot 3$
1	1	0	$1 \cdot 1$
1	1	1	$1 \cdot 2$

Factor multiplication is commutative and associative. It is time and space exponential in the number of variables in the resulting factor.

Factors: Conditioning

X	Y	$f(X, Y)$
0	0	1
0	1	3
1	0	2
1	1	1

Y	$f(0, Y)$
0	1
1	3

Factors: Marginalization

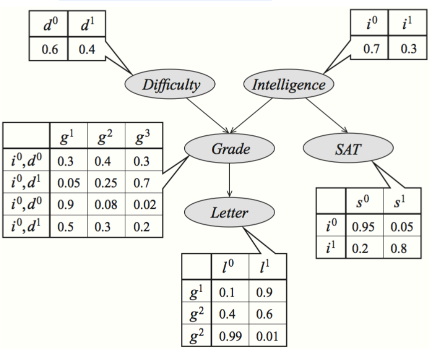
X	Y	$f(X,Y)$
0	0	1
0	1	3
1	0	2
1	1	1

Y	$\sum_{x \in X} f(x,Y)$
0	$1+2$
1	$3+1$

Factor marginalization is commutative. It is time exponential in the number of variables in the original factor, and space exponential in the number of variables in the resulting factor.

Factors: Example

Given the following BN:



compute: $Pr(S \mid I) \times Pr(I)$, $Pr(G \mid i^1, D)$, $\sum_G Pr(L \mid G)$, $\sum_L Pr(L \mid G)$

The basic property: distributivity

When dealing with integers, we know that:

$$\sum_{i=1}^{n} K \times i = K \times \sum_{i=1}^{n} i$$

The same happens when dealing with factors (e.g., $D(X) = \{x^0, x^1\}$):

$$\sum_{X} f(Z) \times g(X, Y) = f(Z) \times g(x^0, Y) + f(Z) \times g(x^1, Y) =$$

$$= f(Z) \times (g(x^0, Y) + g(x^1, Y)) = f(Z) \times \sum_{X} g(X, Y)$$

VE: prior marginal on a chain

We will illustrate VE in the simplest case (a chain with no evidence):

- Boolean variables X_i
- The BN is a chain:

$$P(X_1, X_2, X_3, X_4) = Pr(X_1)Pr(X_2 \mid X_1)Pr(X_3 \mid X_2)Pr(X_4 \mid X_3)$$

- We have no evidence ($E = \emptyset$)
- The only variable of interest is X_4

Thus, we want to compute:

$$Pr(X_4) = \sum_{X_1} \sum_{X_2} \sum_{X_3} Pr(X_1)Pr(X_2 \mid X_1)Pr(X_3 \mid X_2)Pr(X_4 \mid X_3)$$

VE: prior marginal on a chain

Elimination of X_1:

$$Pr(X_4) = \sum_{X_1}\sum_{X_2}\sum_{X_3} Pr(X_1)Pr(X_2 \mid X_1)Pr(X_3 \mid X_2)Pr(X_4 \mid X_3) =$$

$$= \sum_{X_2}\sum_{X_3}\sum_{X_1} \mathrm{Pr}(X_1)Pr(X_2 \mid X_1)Pr(X_3 \mid X_2)Pr(X_4 \mid X_3) =$$

$$= \sum_{X_2}\sum_{X_3} Pr(X_3 \mid X_2)Pr(X_4 \mid X_3)(\sum_{X_1} \mathrm{Pr}(X_1)Pr(X_2 \mid X_1)) =$$

$$= \sum_{X_2}\sum_{X_3} Pr(X_3 \mid X_2)Pr(X_4 \mid X_3)\lambda_1(X_2)$$

where $\lambda_1(X_2) = \sum_{X_1} \mathrm{Pr}(X_1)Pr(X_2 \mid X_1)$

VE: prior marginal on a chain

Elimination of X_2:

$$P(X_4) = \sum_{X_2} \sum_{X_3} Pr(X_3 \mid X_2) Pr(X_4 \mid X_3) \lambda_1(X_2)$$

$$= \sum_{X_3} \sum_{X_2} Pr(X_3 \mid X_2) Pr(X_4 \mid X_3) \lambda_1(X_2) =$$

$$= \sum_{X_3} Pr(X_4 \mid X_3) (\sum_{X_2} Pr(X_3 \mid X_2) \lambda_1(X_2)) =$$

$$= \sum_{X_3} Pr(X_4 \mid X_3) \lambda_2(X_3) =$$

where $\lambda_2(X_3) = \sum_{X_2} Pr(X_3 \mid X_2) \lambda_1(X_2)$

VE: prior marginal on a chain

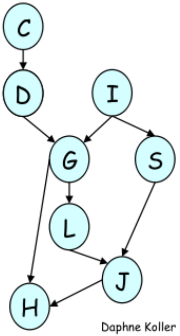
Elimination of X_3:

$$P(X_4) = \sum_{X_3} Pr(X_4 \mid X_3)\lambda_2(X_3) = \lambda_3(X_4)$$

Question:

- What would be the time and space complexity of this execution?

VE: prior marginal on general networks



- **Joint probability distribution:** $P(C, D, I, G, L, S, H, J) = P(H|G, J)P(J|L, S)P(L|G)P(S|I)P(I)P(G|D, I)P(D|C)P(C)$
- **Goal:** $P(G, L, S, H, J)$
- **Elimination order:** C, D, I

VE: prior marginal on general networks

Elimination of C:

$$
\begin{aligned}
& \sum_{I,D,C} P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)P(G|D,I)P(D|C)P(C) = \\
= & \sum_{I,D} P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)P(G|D,I)(\sum_{C} P(D|C)P(C) \\
= & \sum_{I,D} P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)P(G|D,I)\lambda_C(D)
\end{aligned}
$$

VE: prior marginal on general networks

Elimination of D:

$$
\begin{aligned}
& \sum_{I,D} P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)P(G|D,I)\lambda_C(D) \\
= & \sum_{I} P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)(\sum_{D} P(G|D,I)\lambda_C(D)) = \\
= & \sum_{I} P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)\lambda_D(G,I)
\end{aligned}
$$

VE: prior marginal on general networks

Elimination of I:

$$\sum_I P(H|G,J)P(J|L,S)P(L|G)P(S|I)P(I)\lambda_D(G,I)$$

$$P(H|G,J)P(J|L,S)P(L|G)(\sum_I P(S|I)P(I)\lambda_D(G,I)) =$$

$$P(H|G,J)P(J|L,S)P(L|G)\lambda_I(S,G)$$

VE: posterior marginal on general networks

Given the joint probability:

$$P(C, D, I, G, L, S, H, J) =$$
$$P(H|G, J)P(J|L, S)P(L|G)P(S|I)P(I)P(G|D, I)P(D|C)P(C)$$

and evidence $\{I = i, H = h\}$, we want to compute:

$$P(J, S, L \mid i, h) = \frac{P(J, S, L, i, h)}{P(i, h)} = \frac{\sum_{C,D,G} P(C, D, i, G, L, S, h, J)}{\sum_{J,S,L,C,D,G} P(C, D, i, G, L, S, h, J)}$$

using elimination order C, D, G.

Trick

Compute $P(J, S, L, i, h)$ and normalize.

VE: posterior marginal on general networks

Elimination of C:

$$\sum_{C,D,G} P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)P(G|D,i)P(D|C)P(C) =$$

$$\sum_{D,G} P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)P(G|D,i)(\sum_{C} P(D|C)P(C)) =$$

$$\sum_{D,G} P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)P(G|D,i)\lambda_C(D)$$

VE: posterior marginal on general networks

Elimination of D:

$$\sum_{D,G} P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)P(G|D,i)\lambda_C(D) =$$

$$\sum_{G} P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)(\sum_{D} P(G|D,i)\lambda_C(D)) =$$

$$\sum_{G} P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)\lambda_D(G)$$

VE: posterior marginal on general networks

Elimination of G:

$$\sum_G P(h|G,J)P(J|L,S)P(L|G)P(S|i)P(i)\lambda_D(G) =$$

$$P(J|L,S)P(S|i)P(i)(\sum_G P(h|G,J)P(L|G)\lambda_D(G)) =$$

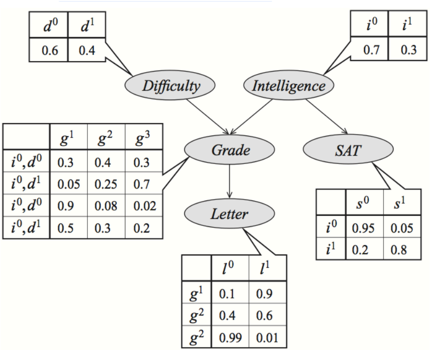
$$P(J|L,S)P(S|i)P(i)\lambda_G(J,L)$$

VE: Summary

- Condition all factors on evidence.
- Eliminate/marginalize each non-query variable X:
  - Move factors mentioning X towards the right.
  - Push-in the summation over X towards the right.
  - Replace expression by λ function.
- Normalize to get distribution.

VE: Exercise

Given the following BN:



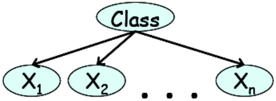
Eliminate Letter and Difficulty from the network.

VE: Some tricks

Given a BN and a set of query nodes Q, and a set of evidence variables E:

- Pruning edges: the effect of conditioning is to remove any edge from a node in E to its parents.
- Pruning nodes: we can remove any leaf node (with its CPT) from the network as long as it does not belong to variables in $Q \cup E$.

Complexity and Elimination Order



$$P(C, X_1, \ldots, X_n) = P(C)P(X_1|C)P(X_2|C)\ldots P(X_n|C)$$

- Eliminate X_n first:

$$\sum_{C, X_1, X_2, \ldots X_{n-1}} P(C)P(X_1|C)P(X_2|C)\ldots P(X_{n-1}|C)(\sum_{X_n} P(X_n|C)) =$$

$$\sum_{C, X_1, X_2, \ldots, X_{n-1}} P(C)P(X_1|C)P(X_2|C)\ldots P(X_{n-1}|C)\lambda_{X_n}(C)$$

- Eliminate C first:

$$\sum_{X_1, X_2, \ldots, X_n} (\sum_{C} P(C)P(X_1|C)P(X_2|C)\ldots P(X_n|C))$$

$$\sum_{X_1, X_2, \ldots, X_n} \lambda_C(X_1, X_2, \ldots, X_n)$$