Inferences in Bayesian Networks
Variable Elimination Algorithm

Alice Gao

Lecture 13

Readings: RN 14.4. PM 8.4.

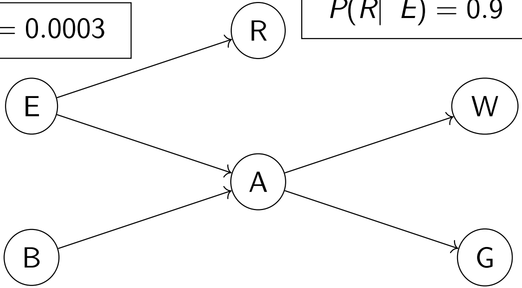
Outline

Learning Goals

By the end of the lecture, you should be able to

- Explain how we can perform probabilistic inference more efficiently using the variable elimination algorithm.
- Define factors. Manipulate factors using operations restrict, sum out, multiply and normalize.
- Describe/trace/implement the variable elimination algorithm for calculating a prior or a posterior probability given a Bayesian network.

A Bayesian Network for the Holmes Scenario



$P(E) = 0.0003$

$P(B) = 0.0001$

$$P(R|\neg E) = 0.0002$$
$$P(R|\ E) = 0.9$$

$$P(W|\neg A) = 0.4$$
$$P(W|\ A) = 0.8$$

$$P(G|\neg A) = 0.04$$
$$P(G|\ A) = 0.4$$

$$P(A|\neg B \wedge \neg E) = 0.01$$
$$P(A|\neg B \wedge\ E) = 0.2$$
$$P(A|\ B \wedge \neg E) = 0.95$$
$$P(A|\ B \wedge\ E) = 0.96$$

Answering a Question

What is the probability that a burglary is happening given that Dr. Watson and Mrs. Gibbon both call?

$$P(B = b | W = t \wedge G = t), b \in \{t, f\}$$

B can be true or false

Shorthand notation:

$$P(B | w \wedge g)$$

b: B is true.

$\neg$b: B is false.

- Query variables: B
- Evidence variables: W and G
- Hidden variables: A, E, and R.

Answering the query

$$\frac{P(B \wedge w \wedge g)}{P(w \wedge g)}$$

$$=$$

$$P(B|w \wedge g) = \frac{P(B \wedge w \wedge g)}{\underbrace{P(b \wedge w \wedge g)}_{1} + \underbrace{P(\neg b \wedge w \wedge g)}_{2}}$$

$$P(B \wedge w \wedge g)$$

$$= \sum_a \sum_e \sum_r P(B \wedge e \wedge a \wedge w \wedge g \wedge r)$$

$$= \sum_a \sum_e \sum_r P(B) P(e) P(r|e) P(a|B \wedge e) P(w|a) P(g|a)$$

$$= P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e) \underbrace{\boxed{\sum_r P(r|e)}}_{=1}$$

Number of operations

How many addition and multiplication operations do we need to perform to evaluate the following expression?

8 times   8 times

$$P(B \wedge w \wedge g) = \sum_a \sum_e \sum_r P(B) P(e) P(r|e) P(a|B \wedge e) P(w|a) P(g|a)$$

1 time   2 times   1   2   3   4   5

(A) ≤ 10

(B) 11-20

(C) 21-40

(D) 41-60

(E) ≥ 61

7 additions.

5 * 8 = 40 multiplications.

47 operations.

Number of operations

How many addition and multiplication operations do we need to perform to evaluate the following expression?

$$P(B \wedge w \wedge g)$$
$$= P(B) \boxed{\sum_a P(w|a) P(g|a) \boxed{\sum_e P(e) P(a|B \wedge e)}}$$

1 multiplication

2 multiplications
1 addition.

10 operations, 1 addition.

(A) ≤ 10

(B) 11-20

(C) 21-40

(D) 41-60

(E) ≥ 61

12 operations.

Introducing the Variable Elimination Algorithm

- Performing probabilistic inference is challenging.
  - Calculating the posterior distribution of one or more query variables given some evidence is #NP.
  - no general efficient implementation available.
- Exact and approximate inferences.
  - naive approach for exact inference: enumerate all the worlds consistent w/ evidence.
  - can do better w/ ↓
- The variable elimination algorithm uses dynamic programming and exploits the conditional independence.

do the calculations once and save the results for later.

VEA = factors + operations on factors
(restrict, sum out, multiply, normalize.)

Factors

- A function from some random variables to a number.
- $f(X_1, \ldots, X_j)$: a factor f on variables $X_1, \ldots, X_j$.
- A factor can represent a joint or a conditional distribution. For example, $f(X_1, X_2)$ can represent $P(X_1 \wedge X_2)$, $P(X_1|X_2)$ or $P(X_1 \wedge X_3 = v_3|X_2)$.
- Define a factor for every conditional probability distribution in the Bayes net.

$$P(B \wedge w \wedge g) = P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e)$$

$P(B)$, $P(E)$, $P(A|B \wedge E)$, $P(R|E)$, $P(W|A)$, $P(G|A)$

$f_1(B)$, $f_2(E)$, $f_3(A,B,E)$, $f_4(R,E)$, $f_5(W,A)$, $f_6(G,A)$.

Restrict a factor

Restrict a factor by assigning a value to the variable in the factor.

- For each observed variable, restrict the factor to the observed value.
- Restricting $f(X_1, X_2, \ldots, X_j)$ to $X_1 = v_1$, produces a new factor $f(X_1 = v_1, X_2, \ldots, X_j)$ on $X_2, \ldots, X_j$. $f'(X_2, \ldots, X_j)$.
- $f(X_1 = v_1, X_2 = v_2, \ldots, X_j = v_j)$ is a number. $f^*()$

$$P(B \wedge w \wedge g) = P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e).$$

Restrict $f_5(W, A)$ to $W = w$. $\Rightarrow f_7(A)$

Restrict $f_6(G, A)$ to $G = g$ $\Rightarrow f_8(A)$

Restrict a factor

$f_1(X, Y, Z)$:

X	Y	Z	val
t	t	t	0.1
t	t	f	0.9
t	f	t	0.2
t	f	f	0.8
f	t	t	0.4
f	t	f	0.6
f	f	t	0.3
f	f	f	0.7

- What is $f_2(Y, Z) = f_1(x, Y, Z)$?
- What is $f_3(Y) = f_2(Y, \neg z)$?
- What is $f_4() = f_3(\neg y)$?

$f_2(Y, Z)$:

Y	Z	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$f_3(Y)$:

Y	val
t	0.9
f	0.8

$f_4()$: 0.8

Sum out a variable

Sum out a variable.

Summing out X_1 with domain $\{v_1, \ldots, v_k\}$ from factor $f(X_1, \ldots, X_j)$, produces a factor on $X_2, \ldots, X_j$ defined by:

$$\left(\sum_{X_1} f\right)(X_2, \ldots, X_j) = f(X_1 = v_1, \ldots, X_j) + \cdots + f(X_1 = v_k, \ldots, X_j)$$

$$P(B \wedge w \wedge g) = P(B) \sum_{a} P(w|a) P(g|a) \sum_{e} P(e) P(a|B \wedge e)$$

Sum out e first, and a next.

Sum out a variable

$f_1(X, Y, Z)$:

X	Y	Z	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$f_2(X, Z)$

X	Z	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

What is $f_2(X, Z) = \sum_Y f_1(X, Y, Z)$?

Multiplying factors

3 cases: f_1 and f_2 have same variables.
f_1 and f_2 have a subset of variables in common.
f_1 and f_2 have no variable in common.

Multiply two factors together.

The **product** of factors $f_1(X, Y)$ and $f_2(Y, Z)$, where Y are the variables in common, is the factor $(f_1 \times f_2)(X, Y, Z)$ defined by:

element-wise multiplication

$$(f_1 \times f_2)(X, Y, Z) = f_1(X, Y) * f_2(Y, Z).$$

$$P(B \wedge w \wedge g) = P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e)$$

Multiplying factors

$\{X, Y\}$

f_1:

X	Y	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$\{Y, Z\}$

f_2:

Y	Z	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

What is $f_1(X, Y) \times f_2(Y, Z)$?

$f_1 \times f_2$

X	Y	Z	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

Normalize a factor

- Convert it to a probability distribution.
- Divide each value by the sum of all the values.

f_1:

Y	val
t	0.2
f	0.6

f_2:

Y	val
t	0.25
f	0.75

Variable elimination algorithm

To compute $P(X_q | X_{o_1} = v_1 \wedge \ldots \wedge X_{o_j} = v_j)$:

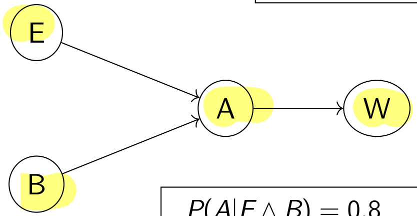
- **Construct a factor** for each conditional probability distribution.
- **Restrict** the observed variables to their observed values.
- Eliminate each hidden variable X_{h_j}. based on some order.
  - **Multiply** all the factors that contain X_{h_j} to get new factor g_j.
  - **Sum out** the variable X_{h_j} from the factor g_j.
- **Multiply** the remaining factors.
- **Normalize** the resulting factor.

Example of VEA

Given a portion of the Holmes network below,
calculate $P(B|\neg A)$ using the variable elimination algorithm.

$P(E) = 0.1$

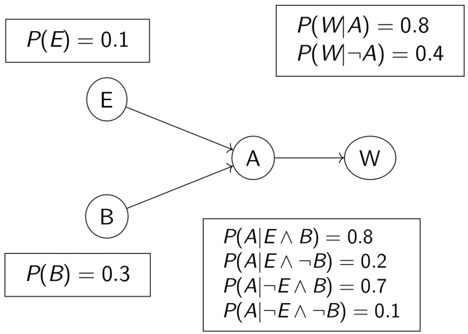
$P(W|A) = 0.8$
$P(W|\neg A) = 0.4$



$P(B) = 0.3$

$P(A|E \wedge B) = 0.8$
$P(A|E \wedge \neg B) = 0.2$
$P(A|\neg E \wedge B) = 0.7$
$P(A|\neg E \wedge \neg B) = 0.1$

Given a portion of the Holmes network below, calculate $P(B|\neg A)$ using the variable elimination algorithm.



$$P(B|\neg a) = \sum_e \sum_w P(B)P(e)P(\neg a|B \wedge e)P(w|\neg a)$$

$$= P(B)\sum_e P(e)P(\neg a|B \wedge e)\sum_w P(w|\neg a)$$

query variable : B

evidence variable : A

hidden variables : E, W

① define factors.

$P(B)$ $P(E)$ $P(A|B \wedge E)$ $P(W|A)$

$f_1(B)$ $f_2(E)$ $f_3(A,B,E)$ $f_4(W,A)$

② restrict factors.

restrict $f_3(A,B,E)$ to $A = \neg a$ to get $f_5(B,E)$

restrict $f_4(W,A)$ to $A = \neg a$ to get $f_6(W)$.

new factor list : $f_1(B)$, $f_2(E)$, $f_5(B,E)$, $f_6(W)$.

③ Get rid of hidden variables E and W.

sum out W first and then E.

(1) multiply all the factors containing W: $f_6(W)$.

sum out W from $f_6(W)$ to get $f_7()$.

(2) multiply all the factors containing E: $f_2(E) \times f_5(B,E) = f_7(B,E)$

sum out E from $f_7(B,E)$ to get $f_8(B)$.

new factor list: $f_1(B)$, $f_8(B)$, $f_7()$.

④ multiply all remaining factors.

$$f_1(B) \times f_8(B) \times f_7() = f_9(B).$$

new factor list: $f_9(B)$.

B	val
t	
f	

⑤ normalize $f_9(B)$ to get $f_{10}(B)$.

QED

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Explain how we can perform probabilistic inference more efficiently using the variable elimination algorithm.
- Define factors. Manipulate factors using operations restrict, sum out, multiply and normalize.
- Describe/trace/implement the variable elimination algorithm for calculating a prior or a posterior probability given a Bayesian network.