CPSC 322: Introduction to Artificial Intelligence

Uncertainty: Variable Elimination Algorithm

Textbook reference: [8.3,8.4]

Instructor: Varada Kolhatkar
University of British Columbia

Credit: These slides are adapted from the slides of the previous offerings of the course. Thanks to all instructors for creating and improving the teaching material and making it available!

Announcements

- Teaching evaluations are open. You should have received an email.
  - I am teaching undergrad for the first time and I will very much appreciate constructive feedback.
- Final exam
  - **Time:** Dec 9 at 7:00pm and **Location:** SRC A
  - Difficulty level: Given that you did so well on midterm we would like to challenge you a bit in the final. So please **start studying now** and make use of all the help available to you.
- Assignment 4 has been released.
  - Due date: **Nov 29th, 11:59 PM**

Lecture outline

- Recap 👈
- Inference in Bayesian networks
- Factors and factor operations
- Variable elimination algorithm
- Variable elimination algorithm examples

Recap: Bayesian networks (BNs) definition

A **Bayesian network** consists of

- A **directed acyclic graph** (V, E) whose nodes are labeled with random variables
- A **domain** for each random variable
- A **conditional probability distribution** for each variable V
  - Specifies $P(V \mid Parents(V))$
  - $Parents(V)$ is the set of variables V' with $(V', V) \in E$. For nodes without predecessors, $Parents(V) = \{\}$

The parents of V are the ones V directly depends upon.
A Bayesian network is a compact representation of the JPD:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

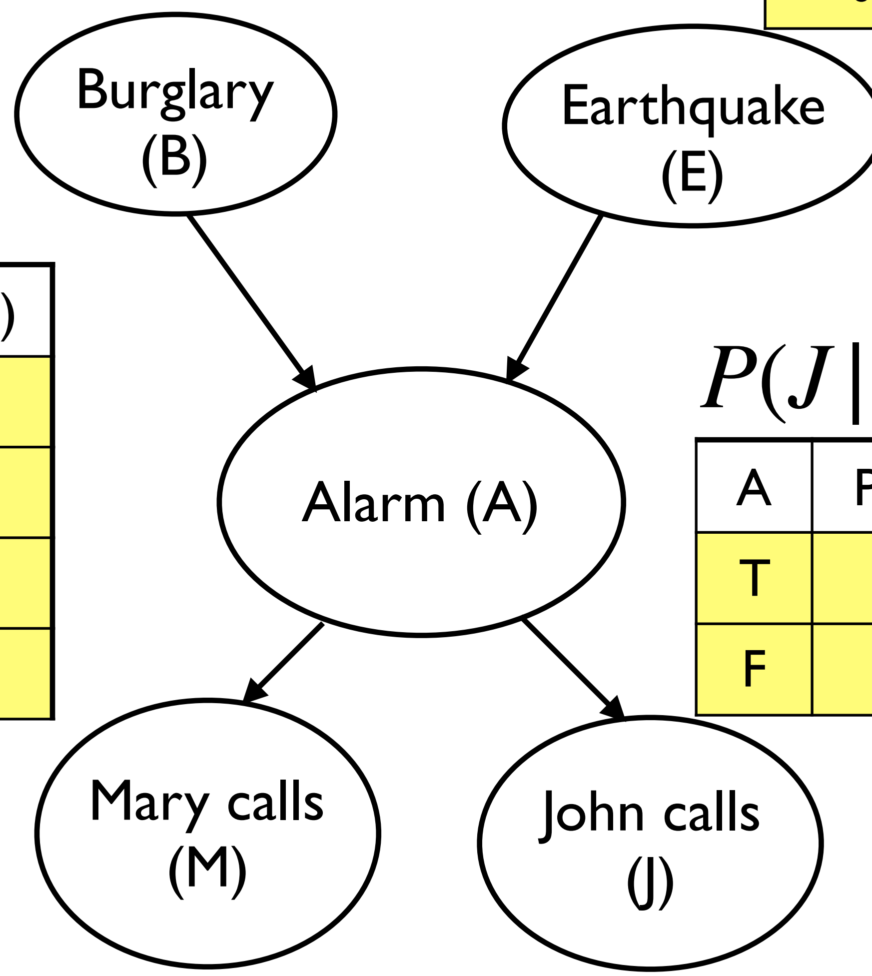
Recap: Bayesian networks

$P(B)$

P(B=T)	P(B=F)
0.001	0.999

$P(E)$

P(E=T)	P(E=F)
0.002	0.998



$P(A \mid B, E)$

B	E	P(A=T|B,E)	P(A=F|B,E)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

$P(J \mid A)$

A	P(J=T|A)	P(J=F|A)
T	0.90	0.10
F	0.05	0.95

$P(M \mid A)$

A	P(M=T|A)	P(M=F|A)
T	0.70	0.30
F	0.01	0.99

Learning outcomes

From this lecture, students are expected to be able to:

- Define factors and apply operations to factors, including assigning, summing out and multiplying factors
- Carry out variable elimination by using factor representation and using the factor operations.
- Use techniques to simplify variable elimination.

Lecture outline

- Recap
- Inference in Bayesian networks
- Factors and factor operations
- Variable elimination algorithm
- Variable elimination algorithm examples

Inference in Bayesian networks

Given:
A Bayesian Network BN
Observations of a subset of its variables $E : E = e$
A subset of its variables Y that is queried

Compute: The conditional probability $P(Y|E = e)$

How: Run **variable elimination algorithm**

Inference in Bayesian networks

Given a belief network, what is the posterior distribution over one (or more) variables, conditioned on one or more observed variables?

<u>Examples:</u>

$$P(Alarm \mid Smoke = F)?$$

$$P(Fire \mid Alarm = T, Leaving = F)?$$

Inference in Bayesian networks

Suppose the variables of the belief network are $X_1, \ldots, X_n$, Z is the query variable, $Y_1 = v_1, \ldots, Y_j = v_j$ are the observed variables (with their values) and $Z_1, \ldots, Z_k$ are the remaining variables, we want to compute $P(Z \mid Y_1 = v_1, \ldots, Y_j = v_j)$

<u>Example:</u>

$P(Leaving \mid Smoke = T, Report = F)$?

$Z = Leaving,$

$Y_1 = ?Y_2 = ?$

$Z_1 = ?, Z_2 = ?, Z_3 = ?$

Inference in Bayesian networks

Suppose the variables of the belief network are $X_1, \ldots, X_n$, Z is the query variable, $Y_1 = v_1, \ldots, Y_j = v_j$ are the observed variables (with their values) and $Z_1, \ldots, Z_k$ are the remaining variables, we want to compute $P(Z \mid Y_1 = v_1, \ldots, Y_j = v_j)$

<u>Example:</u>

$P(Leaving \mid Smoke = T, Report = F)$?

$Z = Leaving,$

$Y_1 = Smoke, Y_2 = Report$

$Z_1 = Tampering, Z_2 = Fire, Z_3 = Alarm$

What do we need to compute?

Remember conditioning and marginalization

$$P(Leaving \mid Smoke = T, Report = F) = \frac{P(Leaving, Smoke = T, Report = F)}{P(Smoking = t, Report = F)}$$

In general,

$$P(Z \mid Y_1 = v_1, \ldots, Y_n = v_n) = \frac{P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}{P(Y_1 = v_1, \ldots, Y_n = v_n)}$$

$$= \frac{P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}{\sum_Z P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}$$

What do we need to compute?

$$P(Z \mid Y_1 = v_1, \ldots, Y_n = v_n) = \frac{P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}{P(Y_1 = v_1, \ldots, Y_n = v_n)}$$

$$= \frac{P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}{\sum_Z P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}$$

- We need to compute the numerator and then normalize

What do we need to compute?

$$\frac{P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}{\sum_Z P(Z, Y_1 = v_1, \ldots, Y_n = v_n)}$$

Note: We can already do all this with Inference by Enumeration. The BN represents the JPD. Could just multiply out the BN to get full JPD and then do Inference by Enumeration BUT that's **extremely inefficient**; it does not scale.

The Variable Elimination (VE) algorithm manipulates conditional probabilities in the form of **factors**.

Lecture outline

- Recap
- Inference in Bayesian networks
- Factors and factor operations 👈
- Variable elimination algorithm
- Variable elimination algorithm examples

Factors

A factor is a function from a tuple of random variables to the real numbers R. We write a factor on variables $X_1, \ldots, X_j$ as $f(X_1, \ldots, X_j)$

A **factor** can denote:

- One distribution
- One *partial* distribution
- Several distributions
- Several *partial* distributions over the given tuple of variables

$P(X_1, X_2)$ is a factor $f(X_1, X_2)$

X_1	X_2	$f(X_1, X_2)$
T	T	0.12
T	F	0.08
F	T	0.08
F	F	0.72

$P(X_1, X_2 = F)$ is a factor $f(X_1)_{X_2=F}$

X_1	X_2	$f(X_1)_{X_2=F}$
T	F	0.08
F	F	0.72

Factors do not have to sum to one.

Factors

A factor is a function from a tuple of random variables to the real numbers R. We write a factor on variables $X_1, \ldots, X_j$ as $f(X_1, \ldots, X_j)$

$P(Z \mid X, Y)$ is a set of probability distributions: one for each combination of values of X and Y

$P(Z = f \mid X, Y)$ is a factor $f(X, Y)_{Z=f}$

$P(X \mid Z, Y)$ is a factor $f(X, Y, Z)$

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

Factors do not have to sum to one.

Operations on factors

A factor is a function from a tuple of random variables to the real numbers R. We write a factor on variables $X_1, \ldots, X_j$ as $f(X_1, \ldots, X_j)$

Operations on factors

- Assigning variables
- Summing out variables
- Multiplication of factors
- Normalizing the factor

Factor operation: assigning a variable

A factor is a function from a tuple of random variables to the real numbers R.

Operation 1: assigning a variable in a factor

E.g., X=t

Factor of Y,X,Z

X	Y	Z	$f_1(X,Y,Z)$
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~~

$$\boxed{f_1(X,Y,Z)_{X=t} = f_2(Y,Z)}$$

Y	Z	$f_2(Y,Z)$
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

Factor of Y,Z

Assignment reduces the factor dimension.

Factor operation: assigning a variable

A factor is a function from a tuple of random variables to the real numbers R.

Operation 1: assigning a variable in a factor

f(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

f(X=t,Y,Z)

Factor of Y,Z

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

f(X=t,Y,Z=f):

Y	val
t	0.9
f	0.8

Factor of Y

f(X=t,Y=f,Z=f): 0.8

Number

Factor operation: sum out a variable

A factor is a function from a tuple of random variables to the real numbers R.
Operation 2: Sum out (marginalize out) a variable from a factor

B	A	C	$f_3(A,B,C)$
t	t	t	0.03
t	t	f	0.07
f	t	t	0.54
f	t	f	0.36
t	f	t	0.06
t	f	f	0.14
f	f	t	0.48
f	f	f	0.32

$$\boxed{\sum_B f_3(A,B,C) = f_4(A,C)}$$

A	C	$f_4(A,C)$
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

Factor operation: multiplying factors

A factor is a function from a tuple of random variables to the real numbers R.

Operation 3: Multiplying factors

$f_1(A,B)$:

A	B	Val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$f_2(B,C)$:

B	C	Val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1(A,B) \times f_2(B,C)$:

A	B	C	val
t	t	t	0.03
t	t	f	0.1x0.7
t	f	t	0.9x0.6
t	f	f	…
f	t	t	
f	t	f	
f	f	t	
f	f	f	

$$f_1(A,B) \times f_2(B,C) = f_3(A,B,C)$$

$$= f_1(A=a, B=b) \times f_2(B=b, C=c) = f_3(A=a, B=b, C=c)$$

Factor operation: normalizing the factor

Divide each entry by the sum of the entries.
The result will sum to 1.

A	$f_8(A)$
t	0.4
f	0.1

A	$f_9(A)$
t	0.4/(0.4+0.1) = 0.8
f	0.1/(0.4+0.1) = 0.2

Summary: Factors and operations on them

A **factor** is a function from a tuple of random variables to the real numbers R.

Operation 1: **assigning** a variable in a factor

E.g., $f_2(Y, Z) = f_1(X, Y, Z)_{X=t}$

Operation 2: **marginalize out** a variable from a factor

E.g., $f_4(A, C) = \sum_B f_3(A, B, C)$

Operation 3: **multiply** two factors

E.g., $f_7(A, B, C) = f_5(A, B) \times f_6(B, C)$

E.g., $f_7(A = a, B = b, C = c) = f_5(A = a, B = b) \times f_6(B = b, C = c)$

Factors and operations on them

iclicker

If we **assign** variable $A = a$ in factor $f_4(A, B)$, what is the correct form for the resulting factor?

A. $f(A)$ B. $f(B)$ ✅ C. $f(A, B)$ D. A number

A **factor** is a function from a tuple of random variables to the real numbers R.

Operation 1: **assigning** a variable in a factor

E.g., $f_2(Y, Z) = f_1(X, Y, Z)_{X=t}$

Operation 2: **marginalize out** a variable from a factor

E.g., $f_4(A, C) = \sum_B f_3(A, B, C)$

Operation 3: **multiply** two factors

E.g., $f_7(A, B, C) = f_5(A, B) \times f_6(B, C)$

E.g., $f_7(A = a, B = b, C = c) = f_5(A = a, B = b) \times f_6(B = b, C = c)$

Factors and operations on them

iclicker

If we **marginalize** out variable B from factor $f_4(A, B)$ what is the correct form of the resulting factor?

A. $f(A)$ ✅ B. $f(B)$ C. $f(A, B)$ D. A number

A **factor** is a function from a tuple of random variables to the real numbers R.

Operation 1: **assigning** a variable in a factor

E.g., $f_2(Y, Z) = f_1(X, Y, Z)_{X=t}$

Operation 2: **marginalize out** a variable from a factor

E.g., $f_4(A, C) = \sum_B f_3(A, B, C)$

Operation 3: **multiply** two factors

E.g., $f_7(A, B, C) = f_5(A, B) \times f_6(B, C)$

E.g., $f_7(A = a, B = b, C = c) = f_5(A = a, B = b) \times f_6(B = b, C = c)$

Factors and operations on them

If we **multiply** factors $f_4(X, Y)$ and $f_5(Z, Y)$, what is the correct form for the resulting factor?

A. $f(X)$ B. $f(X, Z)$ C. $f(X, Y, Z)$ ✅ D. $P(X, Y)$

A **factor** is a function from a tuple of random variables to the real numbers R.

Operation 1: **assigning** a variable in a factor

E.g., $f_2(Y, Z) = f_1(X, Y, Z)_{X=t}$

Operation 2: **marginalize out** a variable from a factor

E.g., $f_4(A, C) = \sum_B f_3(A, B, C)$

Operation 3: **multiply** two factors

E.g., $f_7(A, B, C) = f_5(A, B) \times f_6(B, C)$

E.g., $f_7(A = a, B = b, C = c) = f_5(A = a, B = b) \times f_6(B = b, C = c)$

Factors and operations on them

iclicker

What's the correct form for $\sum_{Y} (f_5(X,Y) \times f_6(Y,Z))$

A. $f(X)$ B. $f(X,Z)$ ✅ C. $f(X,Y,Z)$ D. $f(X,Y)$

A **factor** is a function from a tuple of random variables to the real numbers R.

Operation 1: **assigning** a variable in a factor

E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$

Operation 2: **marginalize out** a variable from a factor

E.g., $f_4(A,C) = \sum_{B} f_3(A,B,C)$

Operation 3: **multiply** two factors

E.g., $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$

E.g., $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

Lecture outline

- Recap
- Inference in Bayesian networks
- Factors and factor operations
- Variable elimination algorithm 👈
- Variable elimination algorithm examples

General inference in Bayesian networks

Given:
A Bayesian Network BN
Observations of a subset of its variables $E : E = e$
A subset of its variables Y that is queried

Compute: The conditional probability $P(Y | E = e)$

Definition of conditional probability

Marginalization over Y: $P(E = e) = \sum_{y' \in dom(Y)} P(Y = y', E = e)$

$$P(Y = y | E = e) = \frac{P(Y = y, E = e)}{P(E = e)} = \frac{P(Y = y, E = e)}{\sum_{y' \in dom Y} P(Y = y', E = e)}$$

All we need to compute is the joint probability of the query variable(s) and the evidence!

Variable Elimination: Intro 1

- We can express the joint probability $P(Y, E_1 = e_1, \ldots, E_j = e_j)$ as a factor $f(Y, E_1, \ldots, E_k, Z_1, \ldots, Z_k)$

- We can compute $P(Y, E_1 = e_1, \ldots, E_j = e_j)$ by

  - **Assigning** $E_1 = e_1, \ldots, E_j = e_j$

  - **Marginalizing** out variables $Z_1, \ldots, Z_k$ one at a time.

  - The order in which we do this is called our **elimination ordering**

$$P(Y, E_1 = e_1, \ldots E_j = e_j) = \sum_{Z_k} \cdots \sum_{Z_1} f(Y, E_1, \ldots, E_k, Z_1, \ldots, Z_k)_{E_1 = e_1, \ldots, E_j = e_j}$$

Variable Elimination: Intro 1

- Are we done?
  - No. This would still represent the whole JPD (as a single factor)
  - We need to exploit the compactness of Bayesian networks

Variable Elimination: Intro 2

- Recall the joint probability distribution of a Bayesian network is:
  $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

- We will have factor f_i for each conditional probability

- For each variable X_i, there is a factor f_i with domain $X_i \cup pa(X_i) : f_1(\{X_i\} \cup pa(X_i)) = P(X_i \mid pa(X_i))$

$$P(Y, E_1 = e_1, \ldots E_j = e_j) = \sum_{Z_k} \cdots \sum_{Z_1} f(Y, E_1, \ldots, E_k, Z_1, \ldots, Z_k)_{E_1=e_1,\ldots,E_j=e_j}$$

$$= \sum_{Z_k} \cdots \sum_{Z_1} \prod_{i=1}^{n} (f_i)_{(E_1=e_1,\ldots,E_j=e_j)}$$

Intuition: Computing sum of products

- Inference in Bayesian networks thus reduces to computing the sums of products
- Example: it takes 9 multiplications to evaluate the expression $ab + ac + ad + aeh + afh + agh$
- How can this expression be evaluated more efficiently?
- Factor out the a and then the h giving $a(b + c + d + h(e + f + g))$
  - This takes only 2 multiplications (same number of additions as above)

Computing sum of products

Similarly how can we compute $\sum_{Z_k} \prod_{i=1}^{n} f_i$ efficiently?

Factor out those terms that do not involve Z_k, e.g.,

$$\sum_{Z_k} f_1(Z_k) f_2(Y) f_3(Z_k, Y) f_4(X, Y)$$

$$= f_2(Y) f_4(X, Y) \Big(\sum_{Z_k} f_1(Z_k) f_3(Z_k, Y) \Big)$$

Summing out a variable efficiently

- To sum out a variable Z from a product $f_1 \times \ldots \times f_k$ of factors:
- Partition the factors into
  those that don't contain Z say $f_1 \times \ldots \times f_i$
  those that contain Z say $f_{i+1} \times \ldots \times f_k$
- We know that $\sum_Z f_1 \times \ldots \times f_k = f_1 \times \ldots \times f_i \times (\sum_Z f_{i+1} \times \ldots \times f_k)$
- We thus have $\sum_Z f_1 \times \ldots \times f_k = f_1 \times \ldots \times f_i \times f'$
- Store f' explicitly and discard $f_{i+1} \ldots f_k$
- Now we have summed out Z

The Variable Elimination algorithm

See the algorithm 8.10 in the text book.

To compute $P(Y = y \mid E = e)$

1. Construct a factor for each conditional probability
2. Assign the observed variables E to their observed values
3. Decompose the sum
4. Sum out all variables $Z_1, \ldots, Z_k$ not involved in the query
5. Multiply the remaining factors (which only involve Y)
6. Normalize by dividing the resulting factor $f(Y)$ by $\sum_{y \in dom(Y)} f(Y)$

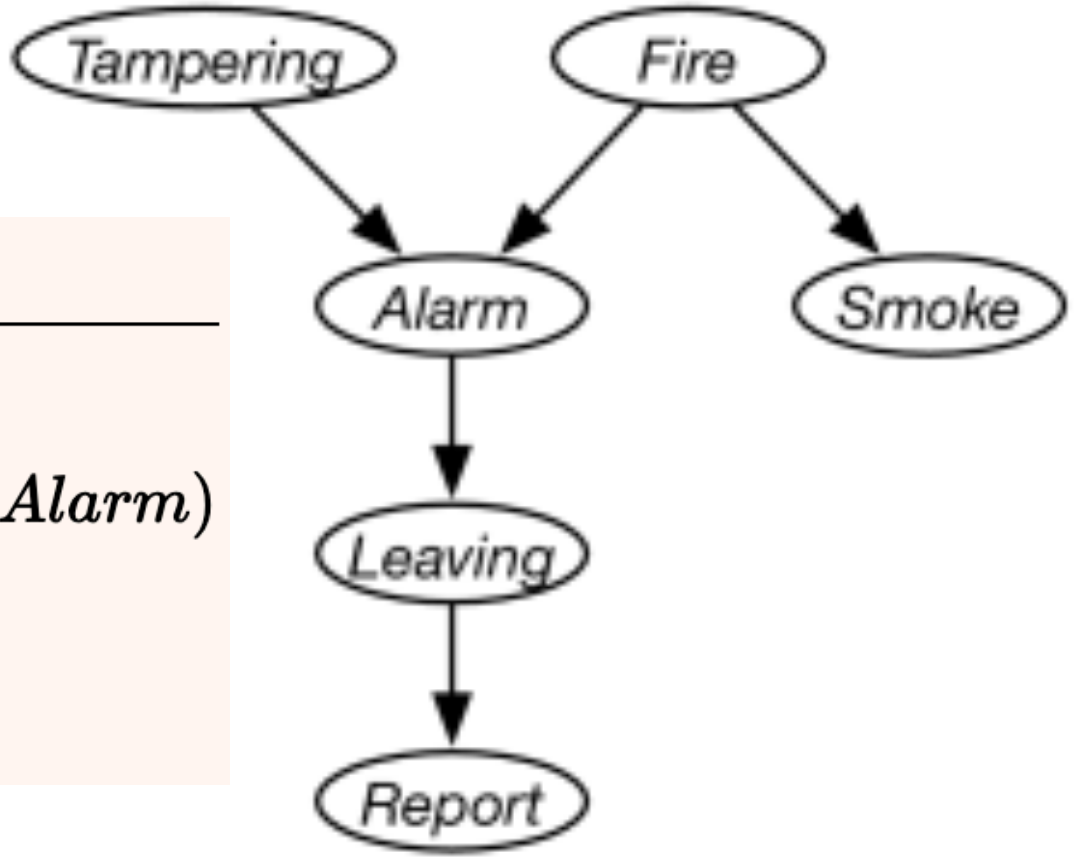
Lecture outline

- Recap
- Inference in Bayesian networks
- Factors and factor operations
- Variable elimination algorithm
- Variable elimination algorithm examples 👈

Fire alarm Example: Textbook

$$P(Tampering \mid Smoke = true \wedge Report = true)$$

$Conditional Probability$	$Factor$
$P(Tampering)$	$f_0(Tampering)$
$P(Fire)$	$f_1(Fire)$
$P(Alarm \mid Tampering, Fire)$	$f_2(Tampering, Fire, Alarm)$
$P(Smoke = yes \mid Fire)$	$f_3(Fire)$
$P(Leaving \mid Alarm)$	$f_4(Alarm, Leaving)$
$P(Report = yes \mid Leaving)$	$f_5(Leaving)$



VE: Fire alarm Example

Find: $P(T \mid S=t, R=t)$

Elimination order: $\boxed{S, R, F, A, L}$

Eliminate the observed variables: S, R

Factors: $f_0(T)$ $\quad$ $f'_3(F)_{S=t} = f_3(F)$

$f_1(F)$ $\quad$ $f_4(A,L)$

$f_2(T,F,A)$ $\quad$ $\cancel{f'_5(R=t)}$ $f'_5(L)_{R=t} = f_5(L)$

Elimination steps for var X

1. Collect all factors containing the variable X
2. Multiply them together
3. Sum out X from the result

Eliminate F

$$\sum_F f_1(F) \times f_2(T,F,A) \times f_3(F)$$

$$= f_7(T,A)$$

Eliminate A

$$\sum_A f_4(A,L) \times f_7(T,A) = f_8(L,T)$$

Eliminate L

$$\sum_L f_5(L) \times f_8(L,T) = f_9(T)$$

$$f_{10}(T) = f_0(T) \times f_9(T)$$

Posterior distribution over T: $\dfrac{f_{10}(T)}{\sum_T f_{10}(T)}$

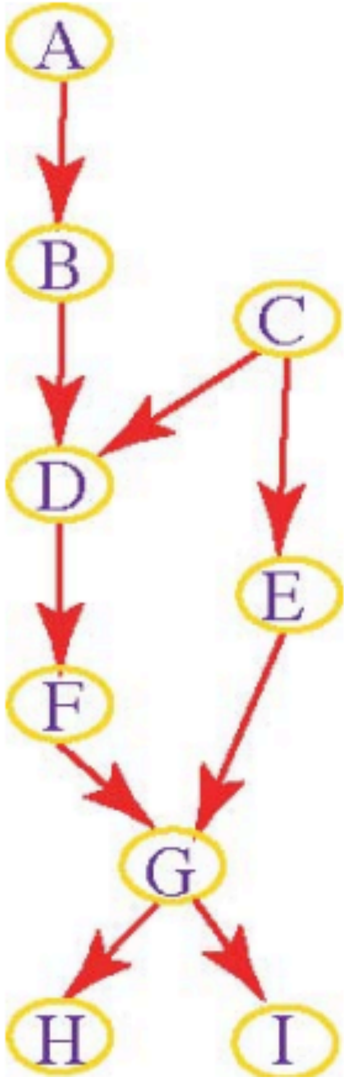
VE Example: Compute $P(G \mid H = h_1)$

Step 1: construct a factor for each conditional probability.

$P(G,H) = \sum_{A,B,C,D,E,F,I} P(A,B,C,D,E,F,G,H,I) =$

$= \sum_{A,B,C,D,E,F,I} P(A)P(B|A)P(C)P(D|B,C)P(E|C)P(F|D)P(G|F,E)P(H|G)P(I|G)$

$= \sum_{A,B,C,D,E,F,I} f_0(A)\ f_1(B,A)\ f_2(C)\ f_3(D,B,C)\ f_4(E,C)\ f_5(F, D)\ f_6(G,F,E)\ f_7(H,G)\ f_8(I,G)$



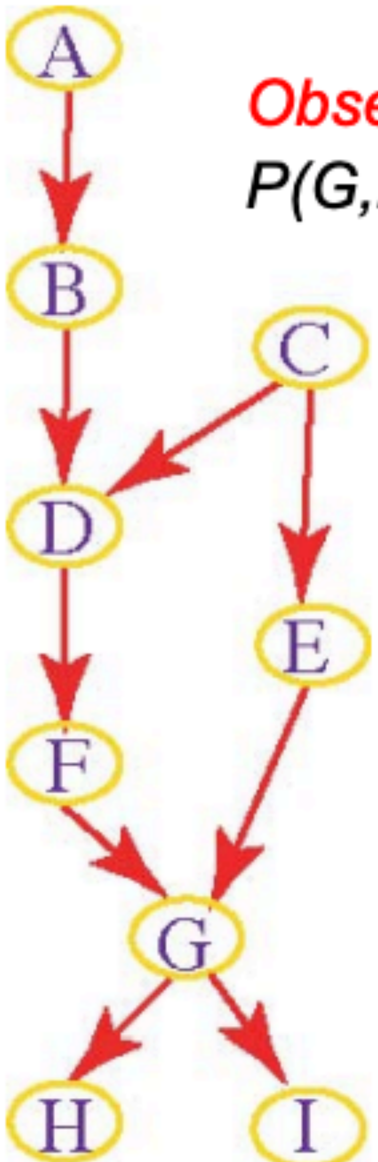
VE Example: Compute $P(G \mid H = h_1)$

Step 2: assign observed variables their observed value

$P(G,H) = \sum_{A,B,C,D,E,F,I} P(A,B,C,D,E,F,G,H,I) =$

$= \sum_{A,B,C,D,E,F,I} P(A)P(B|A)P(C)P(D|B,C)P(E|C)P(F|D)P(G|F,E)P(H|G)P(I|G)$

$= \sum_{A,B,C,D,E,F,I} f_0(A)\ f_1(B,A)\ f_2(C)\ f_3(D,B,C)\ f_4(E,C)\ f_5(F,D)\ f_6(G,F,E)\ f_7(H,G)\ f_8(I,G)$



Observe $H=h_1$:

$P(G,H=h_1)=\sum_{A,B,C,D,E,F,I} f_0(A)\ f_1(B,A)\ f_2(C)\ f_3(D,B,C)\ f_4(E,C)$
$f_5(F,D)\ f_6(G,F,E)\ f_9(G)\ f_8(I,G)$

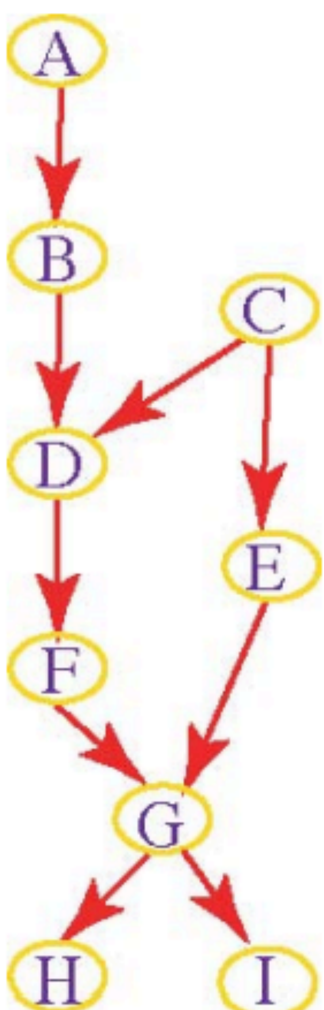
Assigning the variable $H=h_1$:
$f_7(H,G)_{H=h1} = f_9(G)$

VE Example: Compute $P(G \mid H = h_1)$

Step 3: decompose sum

$P(G,H{=}h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= \quad \sum_F \sum_D \quad \sum_B \sum_I \quad \sum_E \quad \sum_C \quad \sum_A$



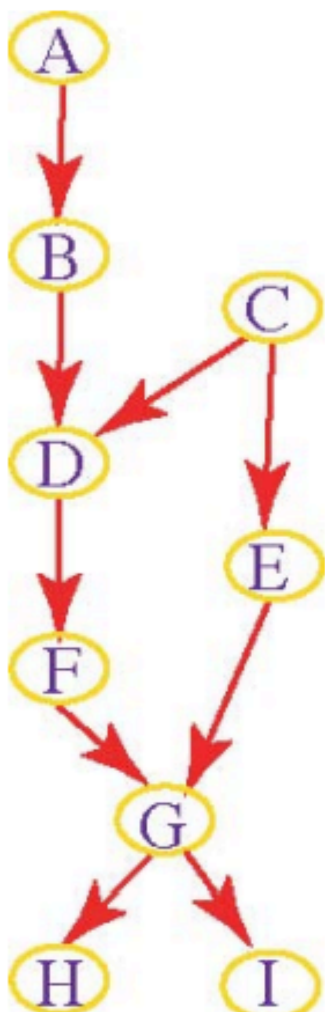
Elimination ordering: A, C, E, I, B, D, F

VE Example: Compute $P(G \mid H = h_1)$

Step 3: decompose sum

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$



Elimination ordering: A, C, E, I, B, D, F

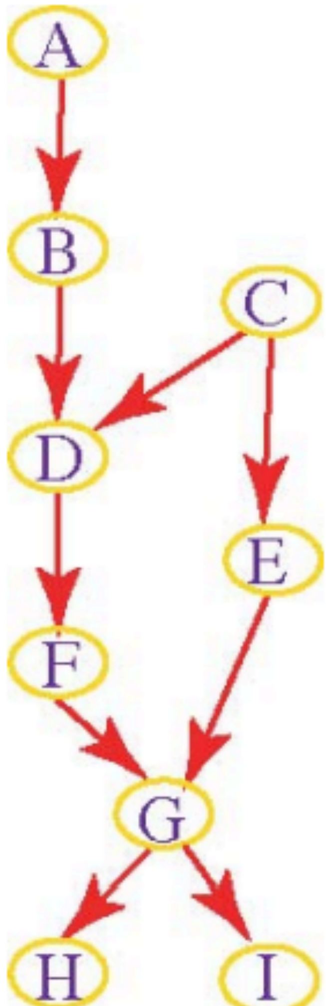
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\ f_1(B,A)\ f_2(C)\ f_3(D,B,C)\ f_4(E,C)\ f_5(F,D)\ f_6(G,F,E)\ f_9(G)\ f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F,D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\ f_3(D,B,C)\ f_4(E,C) \sum_A f_0(A)\ f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F,D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\ f_3(D,B,C)\ f_4(E,C)$



Summing out A: $\sum_A f_0(A)\ f_1(B,A) = f_{10}(B)$
This new factor does not depend on C, E, or I, so we can push it outside of those sums.

Elimination ordering: A, C, E, I, B, D, F

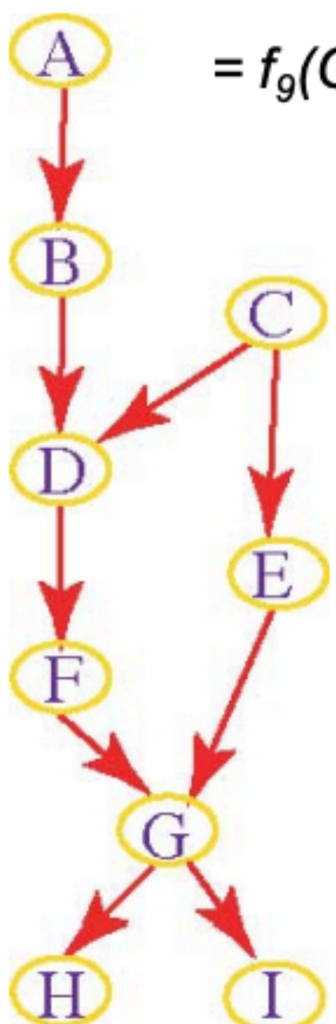
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H{=}h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

Elimination ordering: A, C, E, I, B, D, F

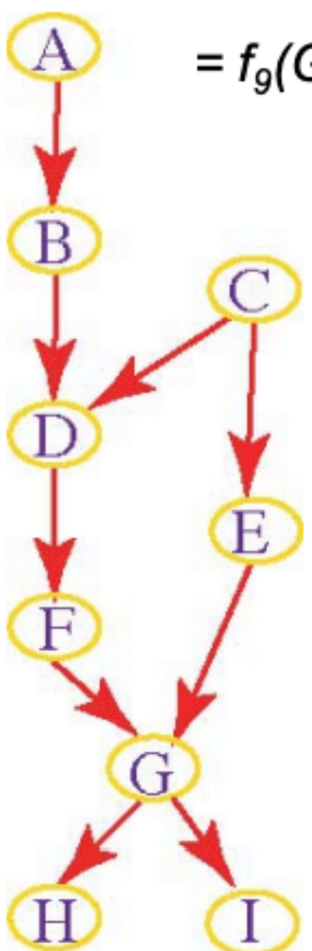
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H{=}h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B) \sum_I f_8(I,G)$

Note the increase in dimensionality:
$f_{12}(G,F,D,B)$ is defined over 4 variables

Elimination ordering: A, C, E, I, B, D, F

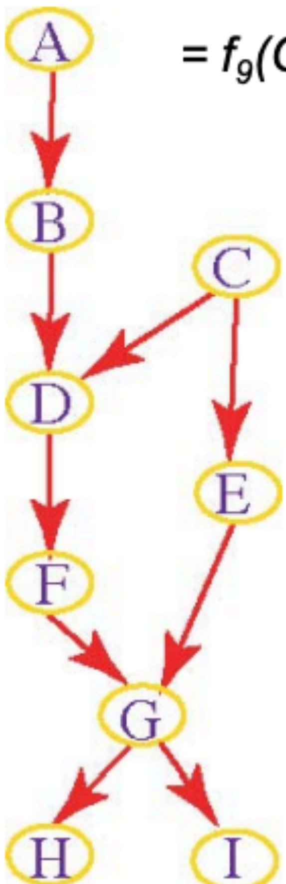
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B) \sum_I f_8(I,G)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B)$

Elimination ordering: A, C, E, I, B, D, F

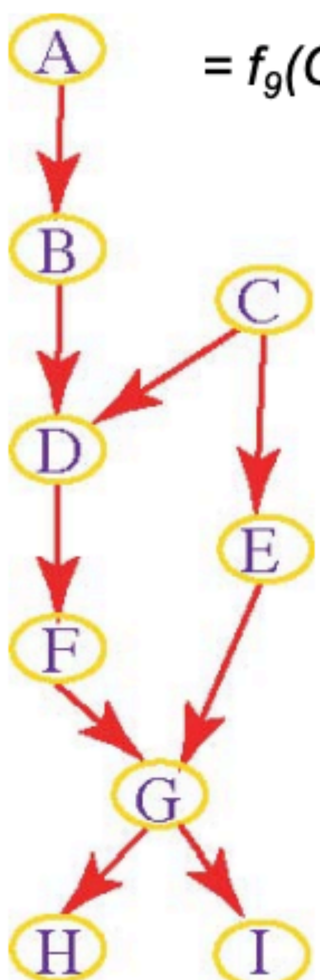
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B) \sum_I f_8(I,G)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D)\, f_{14}(G,F,D)$

Elimination ordering: A, C, E, I, B, D, F

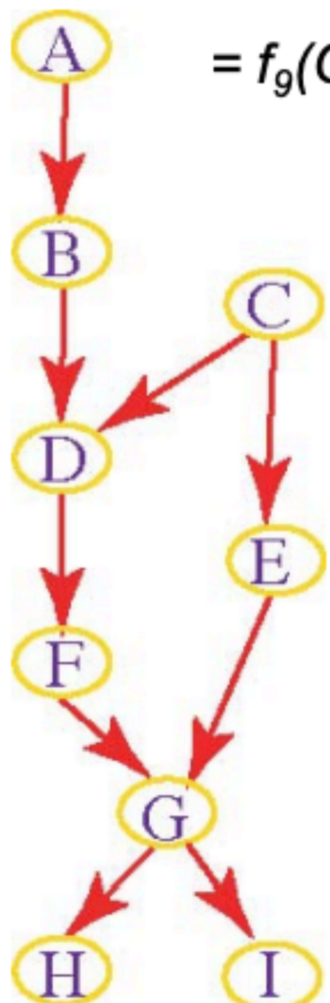
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H{=}h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B) \sum_I f_8(I,G)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D)\, f_{14}(G,F,D)$

$= f_9(G)\, f_{13}(G) \sum_F f_{15}(G,F)$

Elimination ordering: A, C, E, I, B, D, F

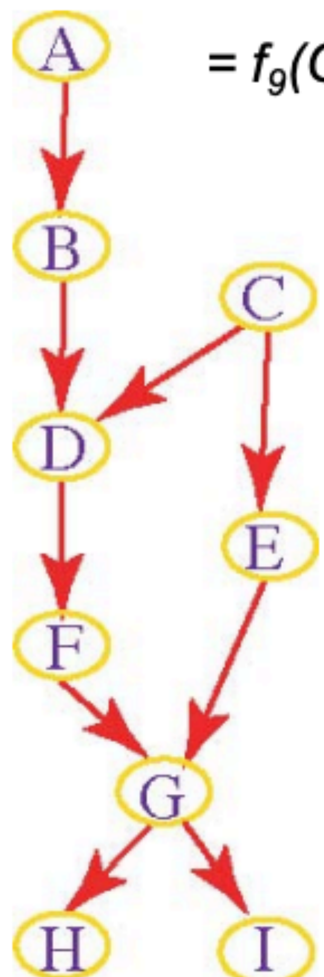
VE Example: Compute $P(G \mid H = h_1)$

Step 4: sum out non- query variables (one at a time)

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B) \sum_I f_8(I,G)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D)\, f_{14}(G,F,D)$

$= f_9(G)\, f_{13}(G) \sum_F f_{15}(G,F)$

$= f_9(G)\, f_{13}(G)\, f_{16}(G)$

Elimination ordering: A, C, E, I, B, D, **F**

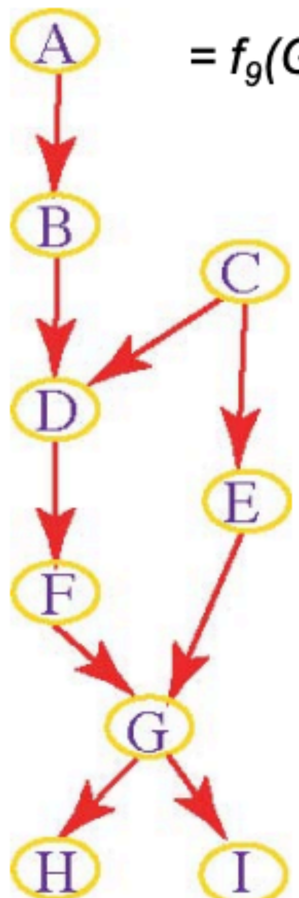
VE Example: Compute $P(G \mid H = h_1)$

Step 5: multiply the remaining factors

$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\ f_1(B,A)\ f_2(C)\ f_3(D,B,C)\ f_4(E,C)\ f_5(F, D)\ f_6(G,F,E)\ f_9(G)\ f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\ f_3(D,B,C)\ f_4(E,C) \sum_A f_0(A)\ f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\ f_3(D,B,C)\ f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\ f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\ f_{12}(G,F,D,B) \sum_I f_8(I,G)$

$= f_9(G)\ f_{13}(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\ f_{12}(G,F,D,B)$

$= f_9(G)\ f_{13}(G) \sum_F \sum_D f_5(F, D)\ f_{14}(G,F,D)$

$= f_9(G)\ f_{13}(G) \sum_F f_{15}(G,F)$

$= f_9(G)\ f_{13}(G)\ f_{16}(G)$

$= f_{17}(G)$

Elimination ordering: A, C, E, I, B, D, F

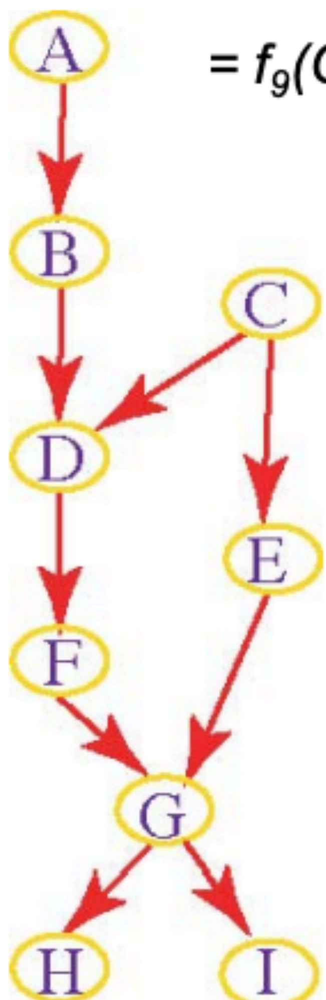
VE Example: Compute $P(G \mid H = h_1)$

Step 6: normalize

$P(G,H{=}h_1) = \sum_{A,B,C,D,E,F,I} f_0(A)\, f_1(B,A)\, f_2(C)\, f_3(D,B,C)\, f_4(E,C)\, f_5(F, D)\, f_6(G,F,E)\, f_9(G)\, f_8(I,G)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C) \sum_A f_0(A)\, f_1(B,A)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E) \sum_C f_2(C)\, f_3(D,B,C)\, f_4(E,C)$



$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B) \sum_I f_8(I,G) \sum_E f_6(G,F,E)\, f_{11}(D,B,E)$

$= f_9(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B) \sum_I f_8(I,G)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D) \sum_B f_{10}(B)\, f_{12}(G,F,D,B)$

$= f_9(G)\, f_{13}(G) \sum_F \sum_D f_5(F, D)\, f_{14}(G,F,D)$

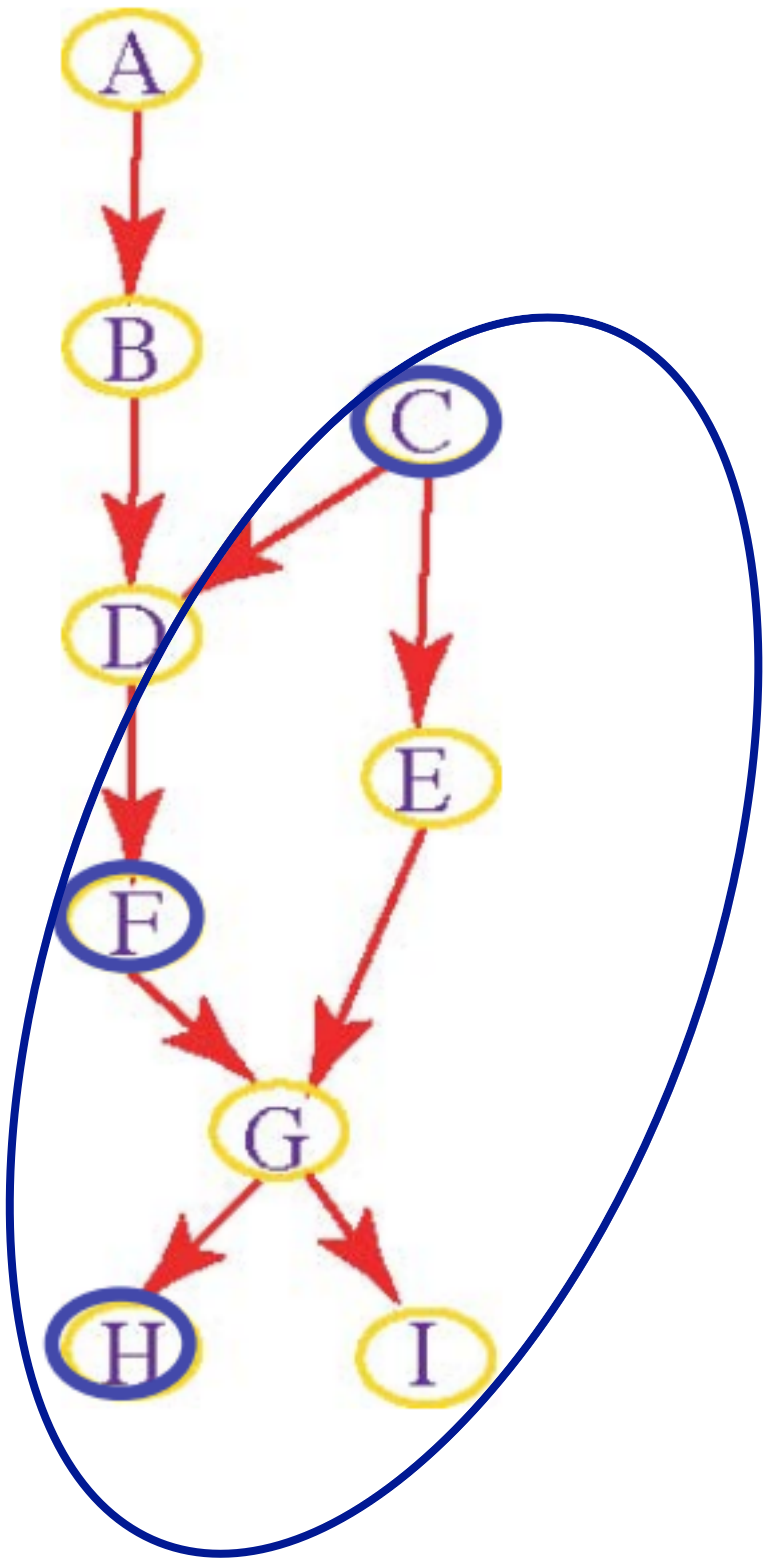
$= f_9(G)\, f_{13}(G) \sum_F f_{15}(G,F)$

$= f_9(G)\, f_{13}(G)\, f_{16}(G)$

$= f_{17}(G)$

$$\boxed{\begin{aligned} P(G = g \mid H = h_1) &= \frac{P(G = g, H = h_1)}{P(H = h_1)} \\ &= \frac{P(G = g, H = h_1)}{\sum_{g' \in dom(G)} P(G = g', H = h_1)} = \frac{f_{17}(g)}{\sum_{g' \in dom(G)} f_{17}(g')} \end{aligned}}$$

VE and conditional independence



Can we use conditional independence to make VE simpler?

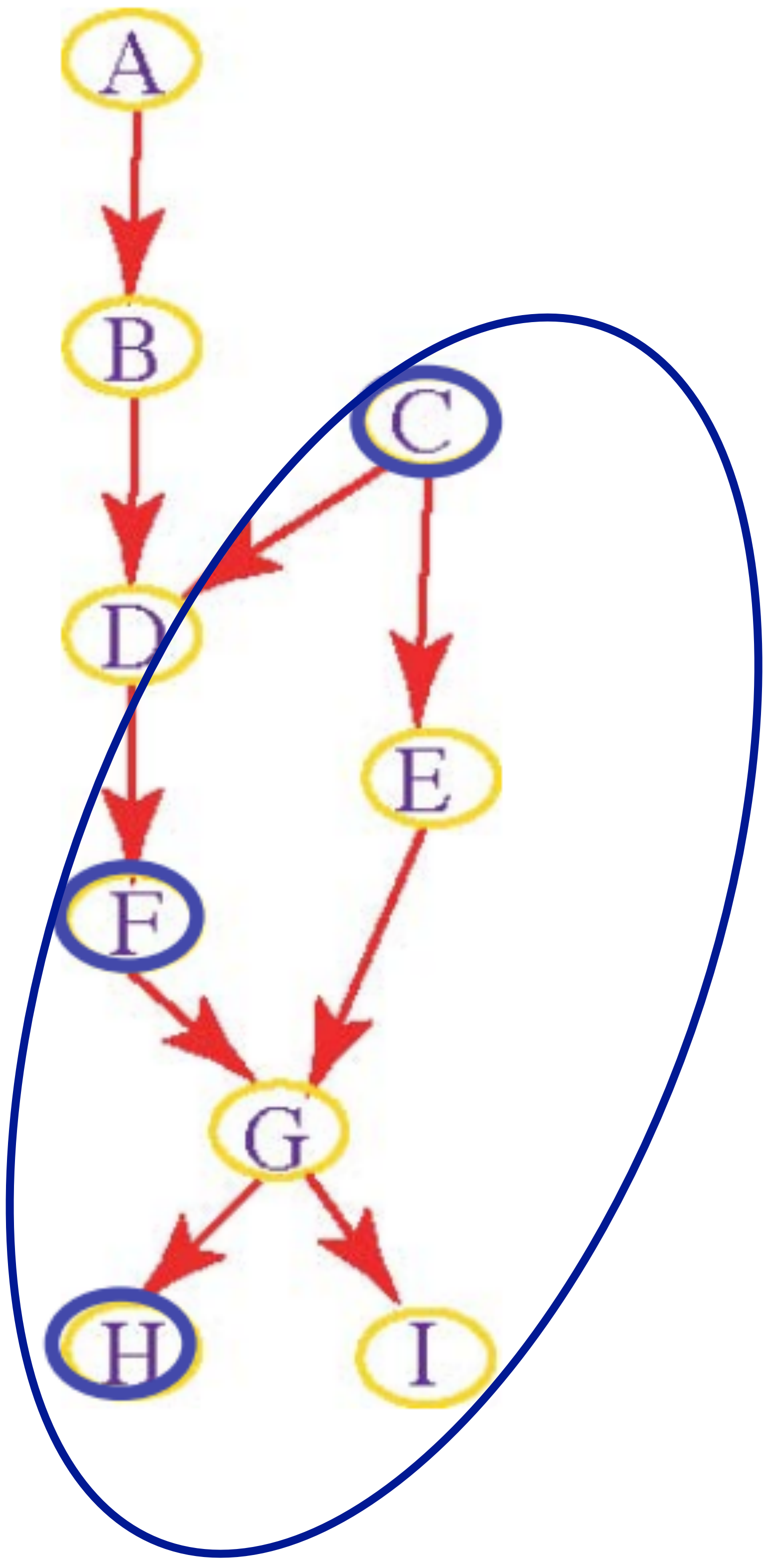
Before running VE, we can prune all variables Z_i that are conditionally independent of the query Y given evidence E: $Z_i \perp\!\!\!\perp Y \mid E$

In particular, any node that has no observed or queried descendants and is itself not observed or queried may be pruned.

Which variables can we prune for the query $P(G = g \mid C = c_1, F = f_1, H = h_1)$?

A, B, D can be pruned.

VE and conditional independence



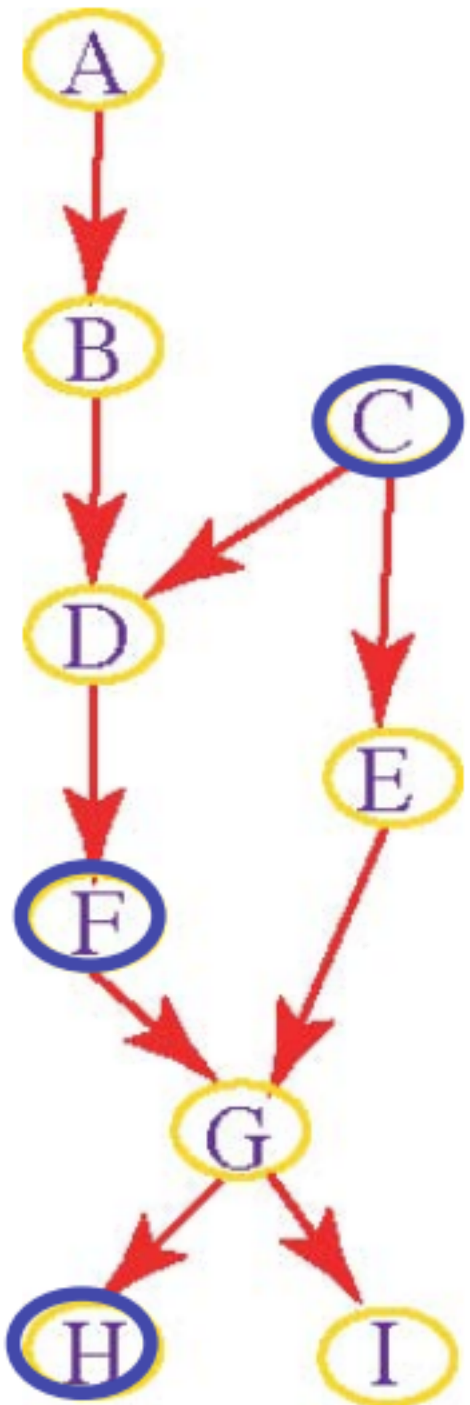
Can we use conditional independence to make VE simpler?

Before running VE, we can prune all variables Z_i that are conditionally independent of the query Y given evidence E: $Z_i \perp\!\!\!\perp Y \mid E$

Which variables can we prune for the query $P(G = g \mid C = c_1, F = f_1, H = h_1)$?

A, B, D can be pruned.

More on pruning



We can also prune unobserved leaf nodes and we can do so recursively.

Example: Which nodes will be pruned if the query is $P(A)$?

- We can recursively prune all nodes except A!

Revisit: Learning outcomes

From this lecture, students are expected to be able to:

- Define factors and apply operations to factors, including assigning, summing out and multiplying factors
- Carry out variable elimination by using factor representation and using the factor operations.
- Use techniques to simplify variable elimination.

Practice exercises

Reminder: they are helpful for staying on top of the material, and for studying for the exam

Exercise 10 is on conditional independence.

Exercise 11 is on variable elimination

Coming up

8.5 Sequential Probability Models