

Main approaches to determine posterior distributions in graphical models:

- Variable Elimination, recursive conditioning: exploit the structure of the network to eliminate (sum out) the non-observed, non-query variables one at a time.

Main approaches to determine posterior distributions in graphical models:

- Variable Elimination, recursive conditioning: exploit the structure of the network to eliminate (sum out) the non-observed, non-query variables one at a time.
- Stochastic simulation: random cases are generated according to the probability distributions.

Main approaches to determine posterior distributions in graphical models:

- Variable Elimination, recursive conditioning: exploit the structure of the network to eliminate (sum out) the non-observed, non-query variables one at a time.
- Stochastic simulation: random cases are generated according to the probability distributions.
- Variational methods: find the closest tractable distribution to the (posterior) distribution we are interested in.

Main approaches to determine posterior distributions in graphical models:

- Variable Elimination, recursive conditioning: exploit the structure of the network to eliminate (sum out) the non-observed, non-query variables one at a time.
- Stochastic simulation: random cases are generated according to the probability distributions.
- Variational methods: find the closest tractable distribution to the (posterior) distribution we are interested in.
- Bounding approaches: bound the conditional probabilities above and below and iteratively reduce the bounds.
- ...

- A **factor** is a representation of a function from a tuple of random variables into a number.
- We write factor f on variables X_1, \dots, X_j as $f(X_1, \dots, X_j)$.

- A **factor** is a representation of a function from a tuple of random variables into a number.
- We write factor f on variables X_1, \dots, X_j as $f(X_1, \dots, X_j)$.
- We can assign some or all of the variables of a factor:
 - ▶ $f(X_1 = v_1, X_2, \dots, X_j)$, where $v_1 \in \text{dom}(X_1)$, is a factor on X_2, \dots, X_j .
 - ▶ $f(X_1 = v_1, X_2 = v_2, \dots, X_j = v_j)$ is a number that is the value of f when each X_i has value v_i .

The former is also written as $f(X_1, X_2, \dots, X_j)_{X_1 = v_1}$, etc.

Example factors

$r(X, Y, Z)$:

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

$r(X=t, Y, Z)$:

Y	Z	val
t	t	0.1
t	f	
f	t	
f	f	

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

Y	Z	val
t	t	0.1
t	f	0.9
f	t	
f	f	

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

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

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

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

Y	val
t	
f	

Example factors

$r(X, Y, Z)$:

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

$r(X=t, Y, Z)$:

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

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

Y	val
t	0.9
f	

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

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

Y	val
t	0.9
f	0.8

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

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

Y	val
t	0.9
f	0.8

$r(X=t, Y=f, Z=f) =$

Example factors

$r(X, Y, Z):$

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

$r(X=t, Y, Z):$

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

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

Y	val
t	0.9
f	0.8

$r(X=t, Y=f, Z=f) = 0.8$

Multiplying factors

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

$$(f_1 * f_2)(\overline{X}, \overline{Y}, \overline{Z}) = f_1(\overline{X}, \overline{Y})f_2(\overline{Y}, \overline{Z}).$$

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	
t	f	t	
t	f	f	
f	t	t	
f	t	f	
f	f	t	
f	f	f	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	
t	f	f	
f	t	t	
f	t	f	
f	f	t	
f	f	f	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	
f	t	t	
f	t	f	
f	f	t	
f	f	f	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	
f	t	f	
f	f	t	
f	f	f	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	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	
f	f	t	
f	f	f	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	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	
f	f	f	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	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	

Multiplying factors example

f_1 :

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

f_2 :

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

$f_1 * f_2$:

A	B	C	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

Summing out variables

We can **sum out** a variable, say X_1 with range $\{v_1, \dots, v_k\}$, from factor $f(X_1, \dots, X_j)$, resulting in a factor on X_2, \dots, X_j defined by:

$$\begin{aligned} & \left(\sum_{X_1} f \right) (X_2, \dots, X_j) \\ &= f(X_1 = v_1, \dots, X_j) + \dots + f(X_1 = v_k, \dots, X_j) \end{aligned}$$

Summing out a variable example

f_3 :

A	B	C	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

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	
f	t	
f	f	

Summing out a variable example

f_3 :

A	B	C	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

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	0.43
f	t	
f	f	

Summing out a variable example

f_3 :

A	B	C	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

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	

Summing out a variable example

f_3 :

A	B	C	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

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

Exercise

Given factors:

s:

A	val
t	0.75
f	0.25

t:

A	B	val
t	t	0.6
t	f	0.4
f	t	0.2
f	f	0.8

o:

A	val
t	0.3
f	0.1

What are the following factors over?

- (a) $s * t$
- (b) $\sum_A s * t$
- (c) $\sum_B s * t$
- (d) $\sum_A \sum_B s * t$
- (e) $s * t * o$
- (f) $\sum_B s * t * o$

- If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \dots \wedge Y_j = v_j$:

$$P(Z | Y_1 = v_1, \dots, Y_j = v_j)$$

=

- If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \dots \wedge Y_j = v_j$:

$$\begin{aligned} &P(Z|Y_1 = v_1, \dots, Y_j = v_j) \\ &= \frac{P(Z, Y_1 = v_1, \dots, Y_j = v_j)}{P(Y_1 = v_1, \dots, Y_j = v_j)} \\ &= \end{aligned}$$

- If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \dots \wedge Y_j = v_j$:

$$\begin{aligned} P(Z | Y_1 = v_1, \dots, Y_j = v_j) \\ &= \frac{P(Z, Y_1 = v_1, \dots, Y_j = v_j)}{P(Y_1 = v_1, \dots, Y_j = v_j)} \\ &= \frac{P(Z, Y_1 = v_1, \dots, Y_j = v_j)}{\sum_Z P(Z, Y_1 = v_1, \dots, Y_j = v_j)}. \end{aligned}$$

- So the computation reduces to the probability of $P(Z, Y_1 = v_1, \dots, Y_j = v_j)$.
- Normalize at the end, by summing out Z and dividing.

Probability of a conjunction

Suppose the variables of the belief network are X_1, \dots, X_n .
To compute $P(Z, Y_1 = v_1, \dots, Y_j = v_j)$, we sum out the other variables, $Z_1, \dots, Z_k = \{X_1, \dots, X_n\} - \{Z\} - \{Y_1, \dots, Y_j\}$.
We order the Z_i into an **elimination ordering**.

$$P(Z, Y_1 = v_1, \dots, Y_j = v_j) \\ =$$

Probability of a conjunction

Suppose the variables of the belief network are X_1, \dots, X_n .
To compute $P(Z, Y_1 = v_1, \dots, Y_j = v_j)$, we sum out the other variables, $Z_1, \dots, Z_k = \{X_1, \dots, X_n\} - \{Z\} - \{Y_1, \dots, Y_j\}$.
We order the Z_i into an **elimination ordering**.

$$\begin{aligned} &P(Z, Y_1 = v_1, \dots, Y_j = v_j) \\ &= \sum_{Z_k} \cdots \sum_{Z_1} P(X_1, \dots, X_n)_{Y_1 = v_1, \dots, Y_j = v_j} \\ &= \end{aligned}$$

Probability of a conjunction

Suppose the variables of the belief network are X_1, \dots, X_n . To compute $P(Z, Y_1 = v_1, \dots, Y_j = v_j)$, we sum out the other variables, $Z_1, \dots, Z_k = \{X_1, \dots, X_n\} - \{Z\} - \{Y_1, \dots, Y_j\}$. We order the Z_i into an **elimination ordering**.

$$\begin{aligned} &P(Z, Y_1 = v_1, \dots, Y_j = v_j) \\ &= \sum_{Z_k} \cdots \sum_{Z_1} P(X_1, \dots, X_n)_{Y_1 = v_1, \dots, Y_j = v_j} \\ &= \sum_{Z_k} \cdots \sum_{Z_1} \prod_{i=1}^n P(X_i | \text{parents}(X_i))_{Y_1 = v_1, \dots, Y_j = v_j} \end{aligned}$$

Computing sums of products

Computation in belief networks reduces to computing the sums of products.

- How can we compute $ab + ac$ efficiently?

Computing sums of products

Computation in belief networks reduces to computing the sums of products.

- How can we compute $ab + ac$ efficiently?
- Distribute out the a giving $a(b + c)$

Computing sums of products

Computation in belief networks reduces to computing the sums of products.

- How can we compute $ab + ac$ efficiently?
- Distribute out the a giving $a(b + c)$
- How can we compute $\sum_{Z_1} \prod_{i=1}^n P(X_i | \text{parents}(X_i))$ efficiently?

Computation in belief networks reduces to computing the sums of products.

- How can we compute $ab + ac$ efficiently?
- Distribute out the a giving $a(b + c)$
- How can we compute $\sum_{Z_1} \prod_{i=1}^n P(X_i | \text{parents}(X_i))$ efficiently?
- Distribute out those factors that don't involve Z_1 .

Variable elimination algorithm

To compute $P(Z|Y_1 = v_1 \wedge \dots \wedge Y_j = v_j)$:

- Construct a factor for each conditional probability.
- Set the observed variables to their observed values.
- Sum out each of the other variables (the $\{Z_1, \dots, Z_k\}$) according to some elimination ordering.
- Multiply the remaining factors. Normalize by dividing the resulting factor $f(Z)$ by $\sum_Z f(Z)$.

Summing out a variable

To sum out a variable Z_j from a product f_1, \dots, f_k of factors:

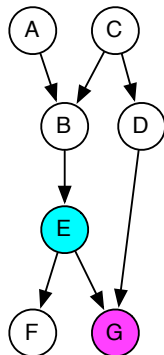
- Partition the factors into
 - ▶ those that don't contain Z_j , say f_1, \dots, f_i ,
 - ▶ those that contain Z_j , say f_{i+1}, \dots, f_k

We know:

$$\sum_{Z_j} f_1 * \dots * f_k = f_1 * \dots * f_i * \left(\sum_{Z_j} f_{i+1} * \dots * f_k \right).$$

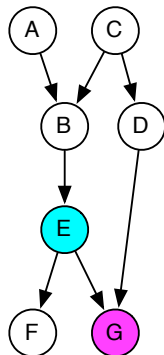
- Explicitly construct a representation of the rightmost factor. Replace the factors f_{i+1}, \dots, f_k by the new factor.

Example



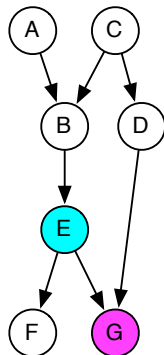
$$P(E \mid g) =$$

Example



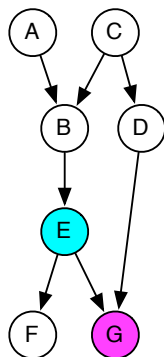
$$P(E \mid g) = \frac{P(E \wedge g)}{\sum_E P(E \wedge g)}$$

Example



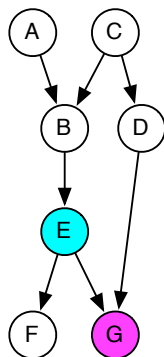
$$\begin{aligned} P(E \mid g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D \end{aligned}$$

Example



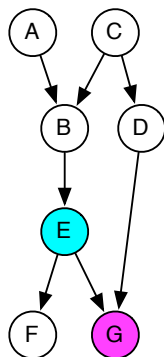
$$\begin{aligned} P(E \mid g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B \mid AC) \\ &\quad P(C)P(D \mid C)P(E \mid B)P(F \mid E)P(g \mid ED) \end{aligned}$$

Example



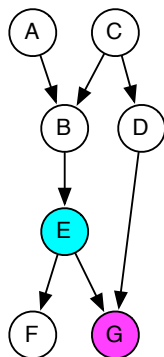
$$\begin{aligned} P(E \mid g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B \mid AC) \\ &\quad P(C)P(D \mid C)P(E \mid B)P(F \mid E)P(g \mid ED) \end{aligned}$$

Example



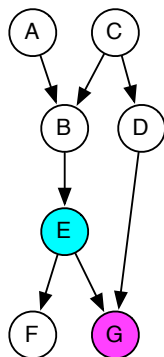
$$\begin{aligned} P(E \mid g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B \mid AC) \\ &\quad P(C)P(D \mid C)P(E \mid B)P(F \mid E)P(g \mid ED) \end{aligned}$$

Example



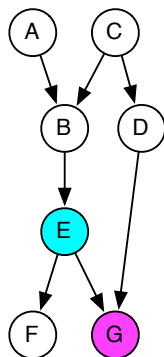
$$\begin{aligned} P(E | g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B | AC) \\ &\quad P(C)P(D | C)P(E | B)P(F | E)P(g | ED) \end{aligned}$$

Example



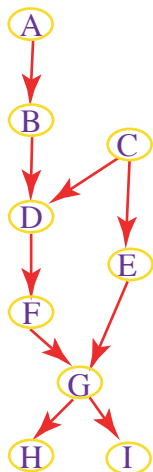
$$\begin{aligned} P(E | g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B | AC) \\ &\quad P(C)P(D | C)P(E | B)P(F | E)P(g | ED) \end{aligned}$$

Example



$$\begin{aligned} P(E \mid g) &= \frac{P(E \wedge g)}{\sum_E P(E \wedge g)} \\ &= \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B \mid AC) \\ &\quad P(C)P(D \mid C)P(E \mid B)P(F \mid E)P(g \mid ED) \end{aligned}$$

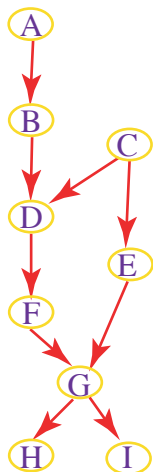
Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A}$$
$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\}$$
$$P(F|D)$$
$$P(G|FE)$$
$$P(H|G) \}$$
$$P(I|G) \}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

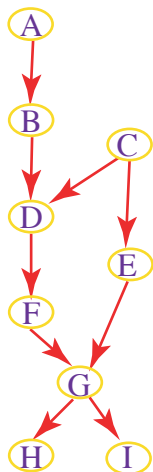
Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A} f_1(B)$$
$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\}$$
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}$$
$$\begin{array}{l} P(H|G) \\ P(I|G) \end{array}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A} f_1(B)$$

$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\} \xrightarrow{\text{elim } C}$$

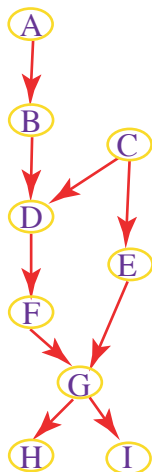
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}$$

$$P(H|G) \}$$

$$P(I|G) \}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

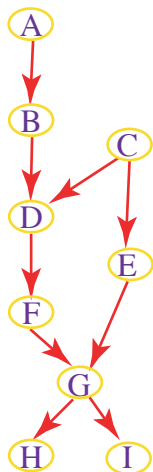
Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A} f_1(B)$$
$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\} \xrightarrow{\text{elim } C} f_2(BDE)$$
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}$$
$$\begin{array}{l} P(H|G) \\ P(I|G) \end{array}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A} f_1(B)$$

$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\} \xrightarrow{\text{elim } C} f_2(BDE)$$

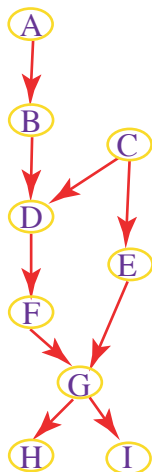
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}$$

$$P(H|G) \} \xrightarrow{\text{obs } H}$$

$$P(I|G) \}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

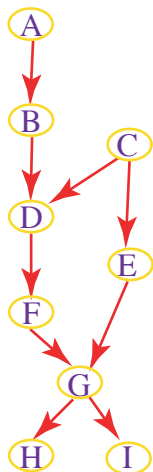
Variable elimination example



$$\begin{array}{l} P(A) \\ P(B|A) \end{array} \left. \vphantom{\begin{array}{l} P(A) \\ P(B|A) \end{array}} \right\} \xrightarrow{\text{elim } A} f_1(B)$$
$$\begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \left. \vphantom{\begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array}} \right\} \xrightarrow{\text{elim } C} f_2(BDE)$$
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array} \left. \vphantom{\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}} \right\} \xrightarrow{\text{obs } H} f_3(G)$$
$$P(H|G) \left. \vphantom{P(H|G)} \right\}$$
$$P(I|G) \left. \vphantom{P(I|G)} \right\}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A} f_1(B)$$

$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\} \xrightarrow{\text{elim } C} f_2(BDE)$$

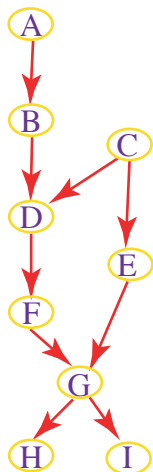
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}$$

$$P(H|G) \} \xrightarrow{\text{obs } H} f_3(G)$$

$$P(I|G) \} \xrightarrow{\text{elim } I}$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

Variable elimination example



$$\left. \begin{array}{l} P(A) \\ P(B|A) \end{array} \right\} \xrightarrow{\text{elim } A} f_1(B)$$

$$\left. \begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array} \right\} \xrightarrow{\text{elim } C} f_2(BDE)$$

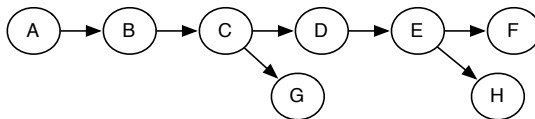
$$\begin{array}{l} P(F|D) \\ P(G|FE) \end{array}$$

$$P(H|G) \} \xrightarrow{\text{obs } H} f_3(G)$$

$$P(I|G) \} \xrightarrow{\text{elim } I} f_4(G)$$

$$P(D, h) = \dots (\sum_A P(A)P(B|A)) (\sum_I P(I|G))$$

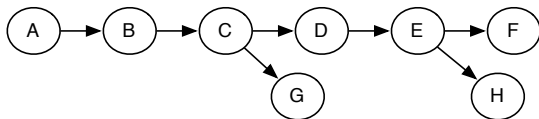
Variable Elimination example



Query: $P(G|f)$; elimination ordering: A, H, E, D, B, C

$$P(G|f) \propto$$

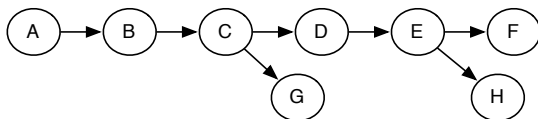
Variable Elimination example



Query: $P(G|f)$; elimination ordering: A, H, E, D, B, C

$$P(G|f) \propto \sum_C \sum_B \sum_D \sum_E \sum_H \sum_A P(A)P(B|A)P(C|B) \\ P(D|C)P(E|D)P(f|E)P(G|C)P(H|E)$$

Variable Elimination example



Query: $P(G|f)$; elimination ordering: A, H, E, D, B, C

$$P(G|f) \propto \sum_C \sum_B \sum_D \sum_E \sum_H \sum_A P(A)P(B|A)P(C|B) \\ P(D|C)P(E|D)P(f|E)P(G|C)P(H|E)$$

$$= \sum_C \left(\sum_B \left(\sum_A P(A)P(B|A) \right) P(C|B) \right) P(G|C) \\ \left(\sum_D P(D|C) \left(\sum_E P(E|D)P(f|E) \sum_H P(H|E) \right) \right)$$

Pruning Irrelevant Variables (Belief networks)

Suppose you want to compute $P(X \mid e_1 \dots e_k)$:

- Prune any variables that have no observed or queried descendants.
- Connect the parents of any observed variable.
- Remove arc directions.
- Remove observed variables.
- Remove any variables not connected to X in the resulting (undirected) graph.