· Exact Inference

want to query the conditional probability $P(\xi \mid E=e)$

$$P(\xi \mid E=e) = \frac{P(\xi, E=e)}{P(E=e)}$$

let $W = \mathcal{X} - \xi - E$ where $\mathcal{X}$ are all variables in the network

$$P(y,e) = \sum_{w} P(y,e,w)$$

$$P(e) = \sum_{y} P(y,e)$$

· Variable Elimination : Intuition

$A \to B \to C \to D$

$$P(B) = \sum_a P(B \mid a) \cdot P(a)$$

$$P(C) = \sum_b P(C \mid b) \cdot P(b)$$

$$P(D) = \sum_c P(D \mid c) \cdot P(c)$$

$X_1 \to X_2 \dashrightarrow X_n$

$$P(X_{i+1}) = \sum_{x_i} P(X_{i+1} \mid x_i) \cdot P(x_i)$$

Suppose each X_i has k values, the CPD $P(X_{i+1} \mid X_i)$ has k^2 values

$P(X_2=1) = \sum_{x_1} P(X_2=1 \mid x_1) \cdot P(x_1)$ sum over k possibilities

$P(X_2=1) \cdots P(X_2=k)$ $\Big\}$ k^2 operations for calculating $P(X_{i+1})$ given $P(X_i)$

$P(X_2)$ $P(X_3)$, $\cdots$ $P(X_n)$ $\Big\}$ $O(nk^2)$

$$P(X_n) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{n-1}} P(x_1) P(x_2 \mid x_1) \cdots P(x_n \mid x_{n-1}) \qquad O(k^n)$$

Using the chain rule, we did the inference over the joint distribution without generating it explicitly

$$
\begin{aligned}
P(d^1) = \; & P(a^1)\ P(b^1 \mid a^1)\ P(c^1 \mid b^1)\ P(d^1 \mid c^1) \\
+ \; & P(a^2)\ P(b^1 \mid a^2)\ P(c^1 \mid b^1)\ P(d^1 \mid c^1) \\
+ \; & P(a^1)\ P(b^2 \mid a^1)\ P(c^1 \mid b^2)\ P(d^1 \mid c^1) \\
+ \; & P(a^2)\ P(b^2 \mid a^2)\ P(c^1 \mid b^2)\ P(d^1 \mid c^1) \\
+ \; & P(a^1)\ P(b^1 \mid a^1)\ P(c^2 \mid b^1)\ P(d^1 \mid c^2) \\
+ \; & P(a^2)\ P(b^1 \mid a^2)\ P(c^2 \mid b^1)\ P(d^1 \mid c^2) \\
+ \; & P(a^1)\ P(b^2 \mid a^1)\ P(c^2 \mid b^2)\ P(d^1 \mid c^2) \\
+ \; & P(a^2)\ P(b^2 \mid a^2)\ P(c^2 \mid b^2)\ P(d^1 \mid c^2)
\end{aligned}
$$

$$
\begin{aligned}
P(d^2) = \; & P(a^1)\ P(b^1 \mid a^1)\ P(c^1 \mid b^1)\ P(d^2 \mid c^1) \\
+ \; & P(a^2)\ P(b^1 \mid a^2)\ P(c^1 \mid b^1)\ P(d^2 \mid c^1) \\
+ \; & P(a^1)\ P(b^2 \mid a^1)\ P(c^1 \mid b^2)\ P(d^2 \mid c^1) \\
+ \; & P(a^2)\ P(b^2 \mid a^2)\ P(c^1 \mid b^2)\ P(d^2 \mid c^1) \\
+ \; & P(a^1)\ P(b^1 \mid a^1)\ P(c^2 \mid b^1)\ P(d^2 \mid c^2) \\
+ \; & P(a^2)\ P(b^1 \mid a^2)\ P(c^2 \mid b^1)\ P(d^2 \mid c^2) \\
+ \; & P(a^1)\ P(b^2 \mid a^1)\ P(c^2 \mid b^2)\ P(d^2 \mid c^2) \\
+ \; & P(a^2)\ P(b^2 \mid a^2)\ P(c^2 \mid b^2)\ P(d^2 \mid c^2)
\end{aligned}
$$

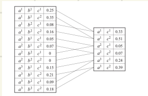
$$P(D) = \sum_C \sum_B \sum_A P(A) \cdot P(B \mid A) \cdot P(C \mid B) \cdot P(D \mid C)$$

$$= \sum_C P(D \mid C) \sum_B P(C \mid B) \cdot \sum_A P(B \mid A)\, P(A)$$

· Variable Elimination

Factor Marginalization:

$\psi(X) = \sum_{Y} \phi(X, Y)$



a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

a^1	c^1	0.33
a^1	c^2	0.51
a^2	c^1	0.05
a^2	c^2	0.07
a^3	c^1	0.24
a^3	c^2	0.39

Figure 9.7 Example of factor marginalization: summing out B.

the factor ψ is the factor marginalization of Y in ϕ, denoted $\sum_Y \phi$

Factor product: $\phi_1 \cdot \phi_2 = \phi_2 \cdot \phi_1$ $\qquad (\phi_1 \cdot \phi_2) \cdot \phi_3 = \phi_1 \cdot (\phi_2 \cdot \phi_3)$

Factor summation: $\sum_X \sum_Y \phi = \sum_Y \sum_X \phi$

if $X \notin Scope[\phi_1]$, then $\phi_1 \cdot \sum_X \phi_2 = \sum_X (\phi_1 \cdot \phi_2)$

$$
\begin{aligned}
P(D) &= \sum_C \sum_B \sum_A P(A, B, C, D) \\
&= \sum_C \sum_B \sum_A \phi_D\, \phi_C\, \phi_B\, \phi_A \\
&= \sum_C \sum_B \phi_D\, \phi_C \sum_A (\phi_B \cdot \phi_A) \\
&= \sum_C \phi_D \sum_B \cdot \phi_C \sum_A \phi_B \cdot \phi_A
\end{aligned}
$$

The form $\sum_Z \prod_{\phi \in \Phi} \phi$ is called "sum-product" inference task

those equal signs are justified by the limited scope of the CPD factors

Let X be some set of variables, Φ be a set of factors such that $\forall \phi \in \Phi$ $Scope[\phi] \subseteq X$

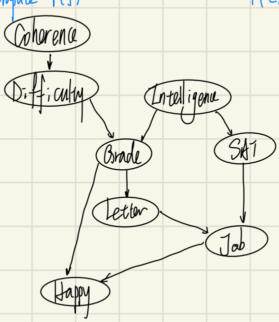
Let $Y \subset X$ be a set of query variables and $Z = X - Y$

$$\phi^*(Y) = \sum_Z \prod_{\phi \in \Phi} \phi$$

for a Bayesian network $\Phi = \{ P(X_i \mid Pa\, X_i) \}$

for a Markov Network ϕ_i is a clique potential

eg. compute $P(J)$



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

$= \phi_C(C) \cdot \phi_D(D, C)\ \phi_I(I)\ \phi_G(G, I, D)\ \phi_S(S, I)\ \phi_L(L|G)\ \phi_J(J, L, S) \cdot \phi_H(H, G, J)$

1° eliminate C

$\tau_1(D) = \sum_C \phi_C(C) \cdot \phi_D(D, C)$ $\quad P(D)$

2° Eliminate (D)

$\tau_2(G, I) = \sum_D \phi_G(G, I, D) \cdot \tau_1(D)$

$\sum_D P(G|I, D) \cdot P(D) = \sum_D P(G, D|I)$

$(I \perp D)$ $\qquad = P(G|I)$

3° eliminate I

$\tau_3(G, S) = \sum_I \phi_I(I)\ \tau_2(G, I) \cdot \phi_S(I, S)$

$P(G, S) = \sum_I P(I) \cdot P(G|I) \cdot P(S|I)$

$(S \perp G \mid I)$

4° eliminate H

$\tau_4(G, J) = \sum_H \phi_H(H, G, J)$

$\sum_H P(H|G, J) = 1$

5° eliminate G

$\tau_5(J, L, S) = \sum_G \tau_4(G, J)\ \phi_L(L, G)\ \tau_3(G, S)$

$P(L, S) = \sum_G P(L|G) \cdot P(G, S)$

$\tau(J) = \sum_{S, L} \tau_5(J, L, S) \cdot \phi_J(J, L, S)$

$\sum_{S, L} P(L, S)\ \ P(J|L, S)$

def eliminate_variable (Φ : set of factors , Z : variable to be eliminated):

$\Phi' = \{\phi \in \Phi : Z \in Scope[\phi]\}$ // factors that contains the target variable

$\Phi'' = \Phi - \Phi''$

$\psi = \prod_{\phi \in \Phi'} \phi$ // factor product

$T = \sum_Z \psi$ // renormalize

return $\Phi'' \cup \{T\}$

def sum_product_variable_elimination(Φ : set of factors , Z : set of variable to be eliminated, $<$: ordering of Z)

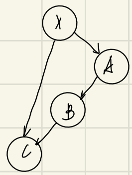
for i = 1, ... k // $Z_i < Z_j$ iff $i<j$

Φ = eliminate_variable(Φ, Z_i)

$\phi^* = \prod_{\phi \in \Phi} \phi$

return ϕ^*

· Variable Elimination : Semantics of Factors



$\Phi = \{ P(X) \ P(A|X) \ P(B|A) \ P(C|B,X) \}$

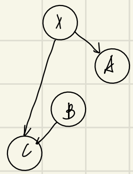
eliminate X

$T(A,B,C) = \sum_X P(X) \ P(A|X) \cdot P(C|B,X)$

T(A,B,C) does not correspond to any probability

1° T(A,B,C) cannot be P(B|...) since P(B|A) is not multiplied

2° $T(A,B,C) = \sum_X P(A,X) \ P(C|B,X) \neq P(A,C|B)$



$\Phi = \{ P(X) , P(A|X) \ P(C|B,X) \}$

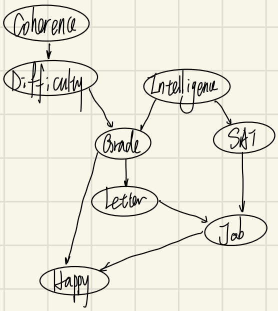
eliminate X

$P(A,B,C) = \sum_X P(X) \ P(B) \ P(A|X) \ P(C|B,X)$

$= P(B) \cdot \sum_X P(X) \ P(A|X) \ P(C|B,X)$

$\therefore \sum_X P(X) \ P(A|X) \ P(C|B,X) = P(A,C|B)$

· Dealing with Evidences



$P(J \mid i^1, h^0)$?

$$P(J \mid i^1, h^0) = \frac{P(J, i^1, h^0)}{P(i^1, h^0)}$$

$$P(C,D,I,G,S,L,J,H) = \phi_C(C)\cdot\phi_D(D,C)\cdot\phi_I(I)\ \phi_G(G,I,D)\cdot\phi_S(S,I)\ \phi_L(L,G)\ \phi_J(J,L,S)\ \phi_H(H,G,J)$$

$$P(C,D,I=i^1,G,S,L,J,H=h^0) = \phi_C(C)\cdot\phi_D(D,C)\ \phi_I[I=i^1]()\cdot\phi_G[I=i^1](G,D)\ \phi_S[I=i^1](S)\ \phi_L(L,G)\cdot\phi_J(J,L,S)\cdot\phi_H[H=h^0](G,J)$$

eliminate C

$$\tau_1(D) = \sum_C \phi_C(C)\cdot\phi_D(D,C)$$

eliminate S

$$\tau_4(J,L) = \sum_S \phi_S[I=i^1](S)\cdot\phi_J(J,L,S)$$

eliminate D

$$\tau_2(G) = \sum_D \tau_1(D)\cdot\phi_I[I=i^1]()\cdot\phi_G[I=i^1](G,D)$$

eliminate G

$$\tau_3(L,J) = \sum_G \tau_2(G)\ \phi_L(L,G)\cdot\phi_H[H=h^0](G,J)$$

$$\phi^*(J) = \sum_L \tau_3(J,L)\cdot\tau_4(J,L) \qquad P(J, I=i^1, H=h^0)$$

$$P(I=i^1, H=h^0) = \sum_j P(J=j, I=i^1, H=h^0)$$

$$P(J \mid I=i^1, H=h^0) = P(J, I=i^1, H=h^0) \,/\, P(I=i^1, H=h^0)$$

def conditional_probability_variable_Elimination(K: A network over $\mathcal{X}$, Y: set of query variables, E=e: Evidence):

    Φ = Factors parameterizing K
    Replace each φ ∈ Φ by φ[E=e]
    Select an elimination ordering <
    Z = 𝒳 - Y - E
    φ* = sum_product_variable_Elimination(Φ, <, E)
    α = Σ_y φ*(y)
    return φ*/α

· Complexity and Graph Structure: Variable elimination

The variable elimination algorithm does not care whether the graph is directed, undirected or partially directed

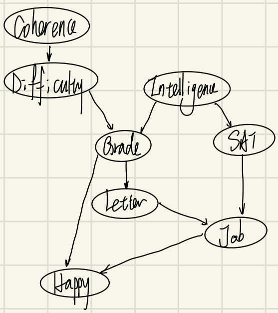
The algorithm's input is a set of factors, the only relevant aspect to the computation is the scope of the factors

Can define the notion of an undirected graph associated with a set of factors

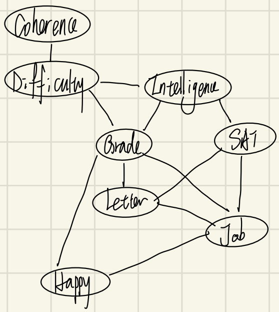
Let Φ be a set of factors. $Scope[\Phi] = \cup\, Scope[\phi] \quad \phi \in \Phi$

define $\mathcal{H}_\Phi$ to be the undirected graph whose nodes correspond to variables in $Scope[\Phi]$,

and $\mathcal{H}_\Phi$ has an edge $X_i - X_j$ iff $\exists\, \phi \in \Phi$ s.t. $\{X_i, X_j\} \subseteq Scope[\Phi]$

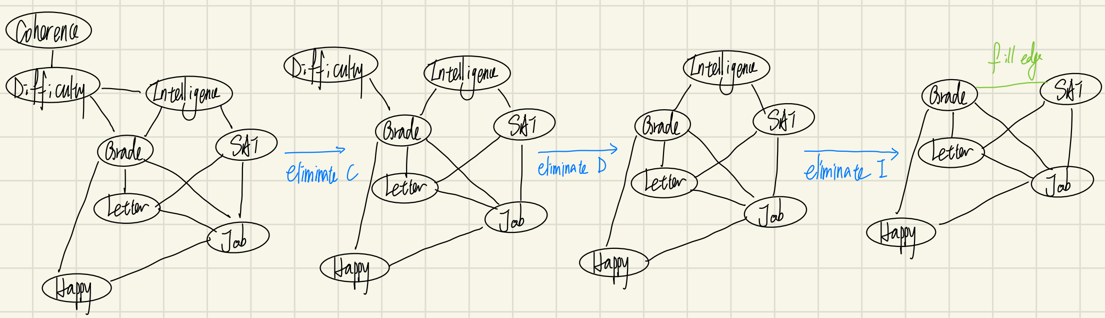


$\Phi = \{ \phi_C(C), \phi_D(D,C), \phi_I(I), \phi_G(D,I), \phi_S(S,I), \phi_L(G,L), \phi_J(J,S), \phi_H(G,J) \}$



Let P be a Gibbs distribution $P(X) = \frac{1}{Z} \prod_{\phi \in \Phi} \phi$, then $\mathcal{H}_\Phi$ is the minimal I-map for P,

and the factors Φ are a parameterization of this network that defines P

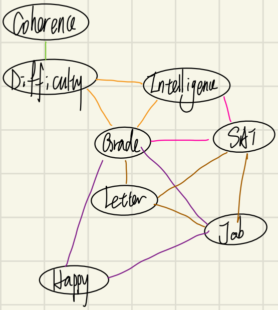


let Φ be a set of factors over $\mathcal{X} = \{X_1, \ldots, X_n\}$ and $<$ be an elimination ordering for some $X \subseteq \mathcal{X}$

The induced graph $I_{\Phi,<}$ is a undirected graph over $\mathcal{X}$, where X_i and X_j are connected if (X_i, X_j) appear some intermediate factor ψ generated by variable elimination using $<$ as an elimination ordering

For a Bayesian Network G, $I_{G,<} = I_{\Phi,<}$ when Φ are CPDs

For a Markov Network $\mathcal{H}$, $I_{\mathcal{H},<} = I_{\Phi,<}$ where Φ are potentials in $\mathcal{H}$



$\tau_1(D) = \sum_C \phi_C(C)\, \phi_D(D,C)$

$\tau_2(G,I) = \sum_D \phi_G(I,D,G)\,,\ \tau_1(D)$

$\tau_3(G,S) = \sum_I \tau_2(G,I)\, \phi_S(S,I)\, \phi_I(I)$

$\tau_4(G,J) = \sum_H \phi_H(G,J)$

$\tau_5(J,L,S) = \sum_G \tau_4(G,J)\, \phi_L(L,G)\, \tau_3(G,S)$

$\tau_6(J,L) = \sum_S \tau_5(J,L,S)\, \phi_J(J,L,S)$

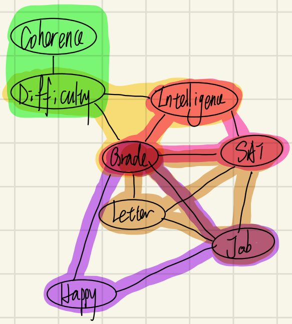
elimination ordering:

C
D
I
H
G
S
L

Let $\mathcal{I}_{\Phi,<}$ be the induced graph for a set of factors Φ and some elimination ordering $<$, then:

The scope of every factor generated during the variable elimination is a clique in $\mathcal{I}_{\Phi,<}$

Every maximal clique in $\mathcal{I}_{\Phi,<}$ is the scope of some intermediate factor in VE



maximal cliques: $\{C,D\}$
$\{D,I,G\}$
$\{I,G,S\}$
$\{G,J,L,S\}$
$\{G,H,J\}$