Exact Inference

Probabilistic Graphical Models

Jerónimo Hernández-González

Concepts review

What we have already seen:

- The probabilistic approach to AI
- What are PGMs? (Representation)
  - Bayesian networks
  - Markov networks
  - Template models (videos)

Now: Exact Inference

- Conditional probability queries
- Variable Elimination
- Message Passing

Conditional probability queries

Definition

Given

- a probability distribution $P(\boldsymbol{X}) = P(X_1, \ldots, X_n)$,
- a partition of $\boldsymbol{X} = (\boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{E})$ into three disjoint subsets of variables, and
- an assignment $\boldsymbol{e}$ to the variables in $\boldsymbol{E}$,

the objective of a *conditional probability query* is to find the probability distribution:

$$P(\boldsymbol{Y}|\boldsymbol{E} = \boldsymbol{e})$$

Complexity

In the general case, this can be rewritten as the following formula:

$$P(\boldsymbol{Y}|\boldsymbol{E}=\boldsymbol{e}) = \sum_{\boldsymbol{h}} P(\boldsymbol{Y}, \boldsymbol{H}=\boldsymbol{h} \mid \boldsymbol{E}=\boldsymbol{e})$$

Exponential complexity!

Complexity

In the general case, this can be rewritten as the following formula:

$$P(\boldsymbol{Y}|\boldsymbol{E}=\boldsymbol{e}) = \sum_{\boldsymbol{h}} P(\boldsymbol{Y}, \boldsymbol{H}=\boldsymbol{h} \mid \boldsymbol{E}=\boldsymbol{e})$$

Exponential complexity!

If P follows a PGM, does complexity reduce?

Factorization

We know that if P factorizes according to a graph, $\mathcal{H}$,

$$P_{\mathcal{H}}(\boldsymbol{X}) = \frac{1}{\Theta} \prod_{i=1}^{f} \phi_i(\boldsymbol{X}_{\phi_i}) = \frac{1}{\Theta} \prod_{i=1}^{f} \phi_i$$

the query can be rewritten as,

$$\begin{aligned} P_{\mathcal{H}}(\boldsymbol{Y} | \boldsymbol{E} = \boldsymbol{e}) &= \sum_{\boldsymbol{h}} P_{\mathcal{H}}(\boldsymbol{Y}, \boldsymbol{H} = \boldsymbol{h} \mid \boldsymbol{E} = \boldsymbol{e}) \\ &= \frac{1}{\Theta} \sum_{\boldsymbol{h}} \prod_{i=1}^{f} \phi_i[\boldsymbol{E} = \boldsymbol{e}] \end{aligned}$$

Non polynomial

In the general case, even using PGMs, the query

$$P_{\mathcal{H}}(\boldsymbol{Y}|\boldsymbol{E}=\boldsymbol{e}) = \sum_{\boldsymbol{h}} P_{\mathcal{H}}(\boldsymbol{Y}, \boldsymbol{H}=\boldsymbol{h} \mid \boldsymbol{E}=\boldsymbol{e})$$
$$= \frac{1}{\Theta} \sum_{\boldsymbol{h}} \prod_{i=1}^{f} \phi_i[\boldsymbol{E}=\boldsymbol{e}]$$

cannot be performed with **exact inference** in polynomial time

Non polynomial

In the general case, even using PGMs, the query

$$P_{\mathcal{H}}(\boldsymbol{Y}|\boldsymbol{E}=\boldsymbol{e}) = \sum_{\boldsymbol{h}} P_{\mathcal{H}}(\boldsymbol{Y}, \boldsymbol{H}=\boldsymbol{h} \mid \boldsymbol{E}=\boldsymbol{e})$$
$$= \frac{1}{\Theta} \sum_{\boldsymbol{h}} \prod_{i=1}^{f} \phi_i[\boldsymbol{E}=\boldsymbol{e}]$$

cannot be performed with **exact inference** in polynomial time

Only for certain specific types of graphs,
exact inference is polynomial

Inference

Answering conditional probability queries

- Variable elimination (a single query)
- Message Passing over clique trees (many queries)

Exact Inference

Probabilistic Graphical Models

Jerónimo Hernández-González

Inference

Answering conditional probability queries

- **Variable elimination (a single query)**
- Message Passing over clique trees (many queries)

Factor algebra

Product

Product of factors

Given two factors ϕ and ψ, their product $\phi \times \psi$ is a new factor whose scope is the union of the scopes of ϕ and ψ ($\Omega_{X_\phi} \cup \Omega_{X_\psi}$) and whose value is the product of ϕ and ψ.

X	Y	ϕ
0	0	3
0	1	2
1	0	4
1	1	1

Y	Z	ψ
0	0	5
0	1	4
0	2	1
1	0	2
1	1	0
1	2	6

X	Y	Z	$\phi \times \psi$
0	0	0	15
0	0	1	12
0	0	2	3
0	1	0	4
0	1	1	0
0	1	2	12
1	0	0	20
1	0	1	16
1	0	2	4
1	1	0	2
1	1	1	0
1	1	2	6

Factor algebra

Reduction

Reduction of a factor

The reduction of a factor ϕ for an assignment of values $\boldsymbol{U} = \boldsymbol{u}$ is a new factor $\phi[\boldsymbol{u}]$ whose scope is $\boldsymbol{V} = \boldsymbol{X}_\phi \setminus \boldsymbol{U}$ and whose value for the assignment $\boldsymbol{V} = \boldsymbol{v}$, $\phi[\boldsymbol{u}](\boldsymbol{v})$, is the value of ϕ for the joint assignment of $\boldsymbol{u}$ and $\boldsymbol{v}$, $\phi[\boldsymbol{u}](\boldsymbol{v}) = \phi(\boldsymbol{u}, \boldsymbol{v})$.

X	Y	Z	ϕ
0	0	0	4
0	0	1	3
0	0	2	5
0	1	0	11
0	1	1	2
0	1	2	1
1	0	0	4
1	0	1	5
1	0	2	12
1	1	0	4
1	1	1	1
1	1	2	9

Y	Z	$\phi[X = 0]$
0	0	4
0	1	3
0	2	5
1	0	11
1	1	2
1	2	1

X	$\phi[Y = 1, Z = 2]$
0	1
1	9

Factor algebra

Marginalization

Marginal

Given a factor ϕ and a set of variables $\boldsymbol{V}$ to remove, the **marginal** $\sum_{\boldsymbol{V}} \phi$ is a factor ψ with scope $\boldsymbol{U} = \boldsymbol{X}_\phi \setminus \boldsymbol{V}$, defined by
$\psi(\boldsymbol{u}) = \sum_{\boldsymbol{v}} \phi(\boldsymbol{u}, \boldsymbol{v})$ **sometimes written as $\phi^{\downarrow \boldsymbol{U}}$

X	Y	Z	ϕ
0	0	0	4
0	0	1	3
0	0	2	5
0	1	0	11
0	1	1	2
0	1	2	1
1	0	0	4
1	0	1	5
1	0	2	12
1	1	0	4
1	1	1	1
1	1	2	9

X	Z	$\sum_Y \phi$
0	0	15
0	1	5
0	2	6
1	0	8
1	1	6
1	2	21

X	$\sum_{Y,Z} \phi$
0	26
1	35

Factor algebra

Normalization

Marginal

Given a factor ϕ, its normalization

$$Norm(\phi)(\boldsymbol{x}) = \frac{1}{\Theta}\phi(\boldsymbol{x})$$

where $Z_\phi = \sum_{\boldsymbol{x}} \phi(\boldsymbol{x})$

Y	Z	ψ	$Norm(\psi)$
0	0	2	2/17
0	1	5	5/17
0	2	2	2/17
1	0	4	4/17
1	1	3	3/17
1	2	1	1/17

Z	ϕ	$Norm(\phi)$
0	8	8/10=0.8
1	2	2/10=0.2

Factor algebra

Relationships of the product

- Product and reduction:

  Let ϕ_1 and ϕ_2 be two factors, and $\boldsymbol{U} = \boldsymbol{u}$ an assignment of values to variables:

  $$(\phi_1 \times \phi_2)[\boldsymbol{U} = \boldsymbol{u}] = \phi_1[\boldsymbol{U} = \boldsymbol{u}] \times \phi_2[\boldsymbol{U} = \boldsymbol{u}]$$

  Apply only to the affected factors

- Product and marginalization:

  Let ϕ_1 and ϕ_2 be two factors, if $X \notin Scope(\phi_1)$:

  $$\sum_X (\phi_1 \times \phi_2) = \phi_1 \times \sum_X \phi_2$$

  Move it towards the affected factors

Exercise

Variable elimination

Given the following MN

$$P(A, B, C, D) = \frac{1}{\Theta}\phi_A(A)\phi_B(B)\phi_C(C)\phi_D(D)$$

where its factors are:

A	ϕ_A
0	4
1	5

B	ϕ_B
0	3
1	2

C	ϕ_C
0	1
1	2

D	ϕ_D
0	6
1	9

Compute:

$$P(D) = \sum_{A,B,C} P(A, B, C, D)$$

Exercise

Variable elimination

Given the following MN

$$P(A, B, C) = \frac{1}{\Theta}\phi_1(A, B)\phi_2(B, C)$$

where its factors are:

A	B	ϕ_1
0	0	2
0	1	6
1	0	1
1	1	4

B	C	ϕ_2
0	0	1
0	1	4
1	0	2
1	1	3

Compute:

$$P(C) = \sum_{A,B} P(A, B, C)$$

Variable elimination

A marginalization problem: sum-product

Problem

Given a distribution $P(\boldsymbol{X})$ that factorizes according to a graph, $\mathcal{H}$,

$$P(\boldsymbol{X}) = \frac{1}{\Theta} \prod_{i=1}^{f} \phi_i(\boldsymbol{X}_i).$$

and $\boldsymbol{Y} \subset \boldsymbol{X}$, assess

$$P(\boldsymbol{Y}) = \sum_{\boldsymbol{z}} P(\boldsymbol{Y}, \boldsymbol{Z}) = \frac{1}{\Theta} \sum_{\boldsymbol{z}} \prod_{i=1}^{f} \phi_i(\boldsymbol{X}_i, \boldsymbol{Z}_i)$$

where $\boldsymbol{Z} = \boldsymbol{X} \backslash \boldsymbol{Y}$

Variable elimination

Problem

Given $P_{\mathcal{H}}(\boldsymbol{X})$ assess $P(\boldsymbol{Y})$ as $Norm(\sum_{\boldsymbol{x}\setminus\boldsymbol{y}} \prod_{i=1}^{f} \phi_i)$ where $\boldsymbol{Y} \subset \boldsymbol{X}$

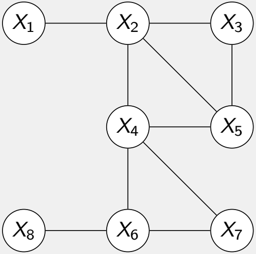
procedure VE($\boldsymbol{\Phi}, \boldsymbol{X}, \boldsymbol{Y}$)
  $\boldsymbol{Z} \leftarrow \boldsymbol{X} \setminus \boldsymbol{Y}$
  for $Z_i \in \boldsymbol{Z}$ **do**
    $\boldsymbol{\Phi} \leftarrow$ Eliminate($\boldsymbol{\Phi}, Z_i$)
  end for
  return $Norm(\prod_{\phi \in \boldsymbol{\Phi}} \phi)$
end procedure

procedure Eliminate($\boldsymbol{\Phi}, V$)
  $\boldsymbol{\Phi}_V \leftarrow \{\phi \in \boldsymbol{\Phi} : V \in Scope(\phi)\}$
  $\boldsymbol{\Phi}_{-V} \leftarrow \boldsymbol{\Phi} \setminus \boldsymbol{\Phi}_V$
  $\psi \leftarrow \prod_{\phi \in \boldsymbol{\Phi}_V} \phi$ ▷ product
  $\tau \leftarrow \sum_V \psi$ ▷ marginalize
  return $\boldsymbol{\Phi}_{-V} \cup \{\tau\}$
end procedure

Variable elimination

Visualizing VE

Assess $P(X_5)$

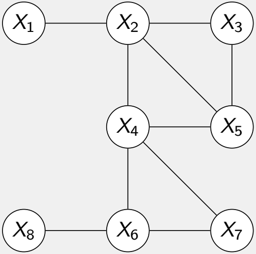


Given $p(\boldsymbol{X}) \propto \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \phi_5(X_4, X_6, X_7) \cdot \phi_6(X_6, X_8)$

Variable elimination

Visualizing VE

Assess $P(X_5)$



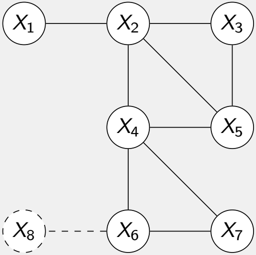
Given $p(\boldsymbol{X}) \propto \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \phi_5(X_4, X_6, X_7) \cdot \phi_6(X_6, X_8)$

$$P(X_5) \propto \sum_{x_1,x_2,x_3,x_4,x_6,x_7,x_8} \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \phi_5(X_4, X_6, X_7) \cdot \phi_6(X_6, X_8)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



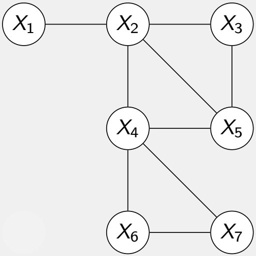
Eliminate X_8

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_4,x_6,x_7} \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \phi_5(X_4, X_6, X_7) \cdot \sum_{x_8} \phi_6(X_6, X_8)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



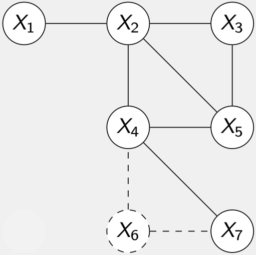
Eliminate X_8

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_4,x_6,x_7} \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \phi_5(X_4, X_6, X_7) \cdot \tau_7(X_6)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



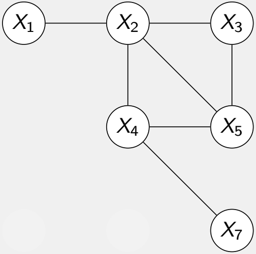
Eliminate X_6

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_4,x_7} \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \sum_{x_6} \phi_5(X_4, X_6, X_7) \times \tau_7(X_6)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



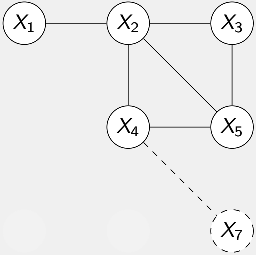
Eliminate X_6

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_4,x_7} \phi_1(X_1,X_2) \cdot \phi_2(X_2,X_3,X_5) \cdot \phi_3(X_2,X_4) \cdot \phi_4(X_4,X_5) \cdot \tau_8(X_4,X_7)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



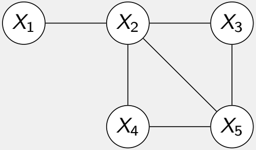
Eliminate X_7

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_4} \phi_1(X_1,X_2)\cdot\phi_2(X_2,X_3,X_5)\cdot\phi_3(X_2,X_4)\cdot \phi_4(X_4,X_5)\cdot \sum_{x_7}\tau_8(X_4,X_7)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



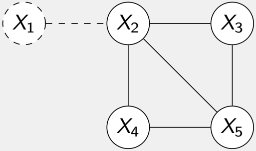
Eliminate X_7

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_4} \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \tau_9(X_4)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



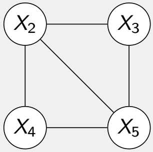
Eliminate X_1

$$p(X_5) \propto \sum_{x_2,x_3,x_4} \phi_2(X_2,X_3,X_5) \cdot \phi_3(X_2,X_4) \cdot \phi_4(X_4,X_5) \cdot \tau_9(X_4) \cdot \sum_{x_1} \phi_1(X_1,X_2)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



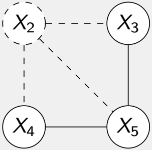
Eliminate X_1

$$p(X_5) \propto \sum_{x_2,x_3,x_4} \phi_2(X_2, X_3, X_5) \cdot \phi_3(X_2, X_4) \cdot \phi_4(X_4, X_5) \cdot \tau_9(X_4) \cdot \tau_{10}(X_2)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



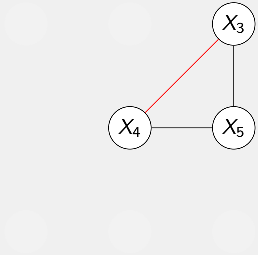
Eliminate X_2

$$p(X_5) \propto \sum_{x_3,x_4} \phi_4(X_4, X_5) \cdot \tau_9(X_4) \cdot \sum_{x_2} \phi_2(X_2, X_3, X_5) \times \phi_3(X_2, X_4) \times \tau_{10}(X_2)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



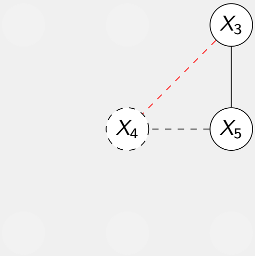
Eliminate X_2

$$p(X_5) \propto \sum_{x_3,x_4} \phi_4(X_4, X_5) \cdot \tau_9(X_4) \cdot \tau_{11}(X_3, X_4, X_5)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



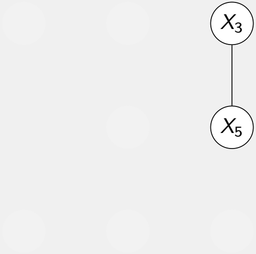
Eliminate X_4

$$p(X_5) \propto \sum_{x_3} \sum_{x_4} \phi_4(X_4, X_5) \times \tau_9(X_4) \times \tau_{11}(X_3, X_4, X_5)$$

Variable elimination

Visualizing VE

Assess $P(X_5)$



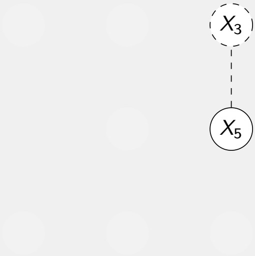
Eliminate X_4

$p(X_5) \propto \sum_{x_3} \tau_{12}(X_3, X_5)$

Variable elimination

Visualizing VE

Assess $P(X_5)$



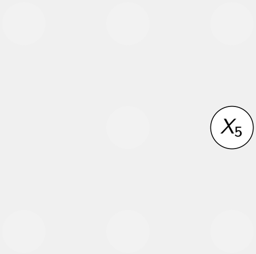
Eliminate X_3

$p(X_5) \propto \sum_{x_3} \tau_{12}(X_3, X_5)$

Variable elimination

Visualizing VE

Assess $P(X_5)$



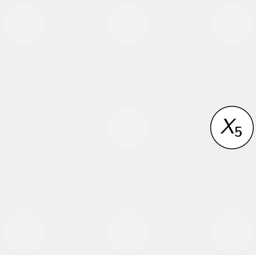
Eliminate X_3

$p(X_5) \propto \tau_{13}(X_5)$

Variable elimination

Visualizing VE

Assess $P(X_5)$



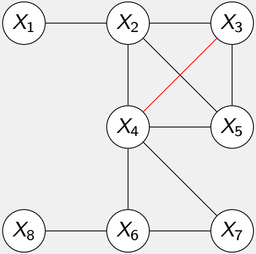
Eliminate X_3

$p(X_5) = Norm[\ \tau_{13}(X_5)\ ]$

Variable elimination

Visualizing VE

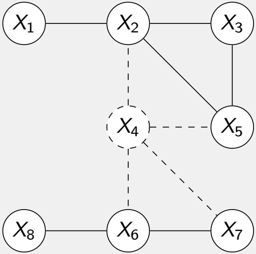
Induced graph for ordering $\{X_8, X_6, X_7, X_1, X_2, X_4, X_3\}$



Variable elimination

A different ordering

Assess $P(X_5)$



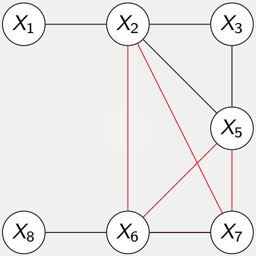
Eliminate X_4

$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_6,x_7,x_8} \phi_1(X_1, X_2) \cdot \phi_2(X_2, X_3, X_5) \cdot \phi_6(X_6, X_8) \cdot \sum_{x_4} \phi_3(X_2, X_4) \times \phi_4(X_4, X_5) \times \phi_5(X_4, X_6, X_7)$$

Variable elimination

A different ordering

Assess $P(X_5)$



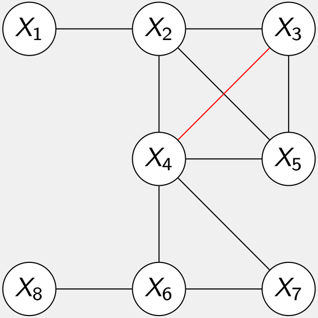
Eliminate X_4

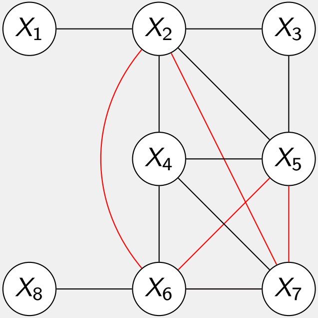
$$p(X_5) \propto \sum_{x_1,x_2,x_3,x_6,x_7,x_8} \phi_1(X_1,X_2)\cdot\phi_2(X_2,X_3,X_5)\cdot\phi_6(X_6,X_8)\cdot \tau_7(X_2,X_5,X_6,X_7)$$

Variable elimination

Visualizing VE

Induced graphs





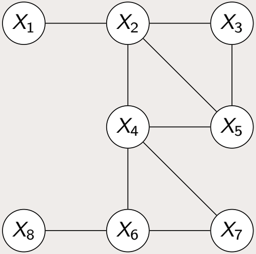
Complexity strongly depends on the order of elimination!

Complexity defined in terms of the size of the largest intermediate factor

Exercise

VE ordering

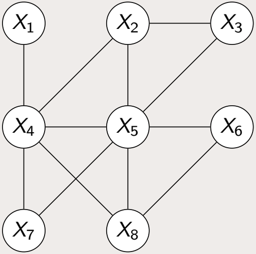
Which is an optimal ordering for assessing $P(X_5)$ in a distribution that factorizes over the following graph?



Exercise

VE ordering

Which is an optimal ordering for assessing $P(X_4)$ in a distribution that factorizes over the following graph?



Variable elimination

Conditional Probability Query with evidence

Problem

Given a distribution $P(\boldsymbol{X})$ that factorizes according to a graph, $\mathcal{H}$,

$$P(\boldsymbol{X}) = \frac{1}{\Theta} \prod_{i=1}^{f} \phi_i(\boldsymbol{X}_i).$$

$\boldsymbol{Y} \subset \boldsymbol{X}$, $\boldsymbol{E} \subset \boldsymbol{X}$, and an assignment $\boldsymbol{e}$ for $\boldsymbol{E}$, assess

$$P(\boldsymbol{Y}|\boldsymbol{E} = \boldsymbol{e}) = \sum_{\boldsymbol{z}} P(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{E} = \boldsymbol{e}) = \frac{1}{\Theta} \sum_{\boldsymbol{z}} \prod_{i=1}^{f} \phi_i(\boldsymbol{Z}_i, \boldsymbol{Y}_i, \boldsymbol{E}_i = \boldsymbol{e}_i)$$

where $\boldsymbol{Z} = \boldsymbol{X} \setminus \boldsymbol{Y} \setminus \boldsymbol{E}$

Variable elimination

Conditional Probability Query with evidence

Problem

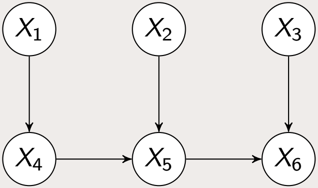
Given $P_{\mathcal{H}}(\boldsymbol{X})$ assess $P(\boldsymbol{Y}|\boldsymbol{E}=\boldsymbol{e})$ as $Norm(\sum_{\boldsymbol{z}} \prod_{i=1}^{f} \phi_i[\boldsymbol{E}=\boldsymbol{e}])$ where $\boldsymbol{Y} \subset \boldsymbol{X}$, $\boldsymbol{E} \subset \boldsymbol{X}$ and $\boldsymbol{Y} \cup \boldsymbol{E} = \emptyset$

procedure Evidence-VE($\Phi, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{E}=\boldsymbol{e}$)
    $\Phi \leftarrow \{\phi[\boldsymbol{E}=\boldsymbol{e}], \forall \phi \in \Phi\}$
    $\phi \leftarrow$ VE($\Phi, \boldsymbol{X} \setminus \boldsymbol{E}, \boldsymbol{Y}$)
    return ϕ
end procedure

Exercise

Intermediate Factors

Consider running variable elimination on this Bayesian network.



Which of the nodes, if eliminated first, results in the largest intermediate factor?

Exercise

Uses of Variable Elimination

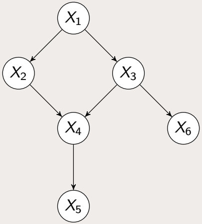
Which of the following quantities can be computed using the sum-product variable elimination algorithm?

a) $p(X|E = e)$ in a Bayesian network
b) $p(X)$ in a Bayesian network
c) The most likely assignment to the variables in a Markov network.
d) The partition function for a Markov network

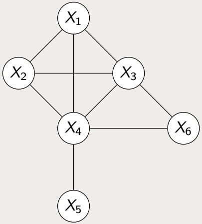
Exercise

Induced graphs

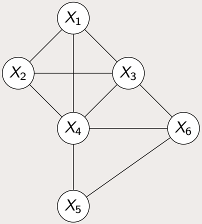
If we perform variable elimination in the BN in the left with the variable ordering $X_2, X_1, X_3, X_6, X_5, X_4$, which is the induced graph?



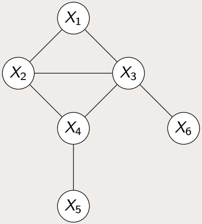
BN



Induced G.1



Induced G.2



Induced G.3

Variable elimination for MAP

Max-Sum

Problem

Given $P_{\mathcal{H}}(\boldsymbol{X})$ assess $\arg\max_{\boldsymbol{z}} P(\boldsymbol{z})$ as

$$\arg\max_{\boldsymbol{z}} \prod_{i=1}^{f} \phi_i[\boldsymbol{X} = \boldsymbol{x}] = \arg\max_{\boldsymbol{z}} \sum_{i=1}^{f} \log \phi_i[\boldsymbol{X} = \boldsymbol{x}]$$

procedure VE-max($\boldsymbol{\Phi}, \boldsymbol{X}$)
    $\log \boldsymbol{\Phi}$
    for $i \in \{1, \ldots, |\boldsymbol{X}|\}$ **do**
        $\boldsymbol{\Phi}; \psi_i \leftarrow$ Elim-Max($\boldsymbol{\Phi}, X_i$)
    end for
    $\boldsymbol{x} \leftarrow$ findMAP($\{\psi_i\}$)
    return $\boldsymbol{x}$
end procedure

procedure Elim-Max($\boldsymbol{\Phi}, V$)
    $\boldsymbol{\Phi}_V \leftarrow \{\phi \in \boldsymbol{\Phi} : V \in Scope(\phi)\}$
    $\boldsymbol{\Phi}_{-V} \leftarrow \boldsymbol{\Phi} \setminus \boldsymbol{\Phi}_V$
    $\psi \leftarrow \sum_{\phi \in \boldsymbol{\Phi}_V} \phi$ ▷ Sum
    $\tau \leftarrow \max_V \psi$ ▷ Max-marg.
    return $\boldsymbol{\Phi}_{-V} \cup \{\tau\}; \psi$
end procedure

procedure findMAP($\{\psi_i\}$)
    for $i \in \{|\boldsymbol{X}|, \ldots, 1\}$ **do**
        $\boldsymbol{u} \leftarrow \boldsymbol{x}^* < Scope(\psi_i) \setminus X_i >$
        $x_i^* \leftarrow \arg\max_{x_i} \psi_i(x_i, \boldsymbol{u})$
    end for
    return $\boldsymbol{x}^*$
end procedure

Factor algebra

Sum

Sum of factors

Given two factors ϕ and ψ, their sum $\phi + \psi$ is a new factor whose scope is the union of the scopes of ϕ and ψ ($\Omega_{X_\phi} \cup \Omega_{X_\psi}$) and whose value is the sum of ϕ and ψ.

X	Y	ϕ
0	0	3
0	1	2
1	0	4
1	1	1

Y	Z	ψ
0	0	5
0	1	4
0	2	1
1	0	2
1	1	0
1	2	6

X	Y	Z	$\phi + \psi$
0	0	0	8
0	0	1	7
0	0	2	4
0	1	0	4
0	1	1	2
0	1	2	8
1	0	0	9
1	0	1	8
1	0	2	5
1	1	0	3
1	1	1	1
1	1	2	7

Factor algebra

Max-Marginalization

Max-marginal

Given a factor ϕ and a set of variables $\boldsymbol{V}$ to remove, the **max-marginal** $\text{máx}_{\boldsymbol{V}} \phi$ is a factor ψ with scope $\boldsymbol{U} = \boldsymbol{X}_\phi \setminus \boldsymbol{V}$, defined by $\psi(\boldsymbol{u}) = \text{máx}_{\boldsymbol{v}} \phi(\boldsymbol{u}, \boldsymbol{v})$

X	Y	Z	ϕ
0	0	0	4
0	0	1	3
0	0	2	5
0	1	0	11
0	1	1	2
0	1	2	1
1	0	0	4
1	0	1	5
1	0	2	12
1	1	0	4
1	1	1	1
1	1	2	9

X	Z	$\text{máx}_Y \phi$
0	0	11
0	1	3
0	2	5
1	0	4
1	1	5
1	2	12

X	$\text{máx}_{Y,Z} \phi$
0	11
1	12

Variable elimination

Summary

- Variable elimination marginalizes
- Answers conditional probability queries too
- It takes as input:
    1. PGM, $\mathcal{H}$
    2. Variables to marginalize out
    3. -optional- Observed variables
    4. Elimination ordering
- Its complexity is exponential in the width of the graph induced by the elimination ordering chosen

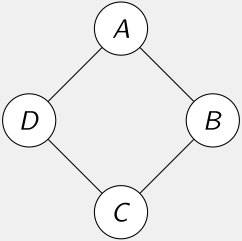
Exact Inference

Probabilistic Graphical Models

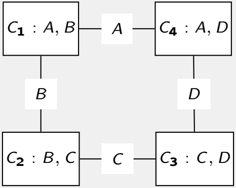
Jerónimo Hernández-González

Belief propagation: message passing

Markov network

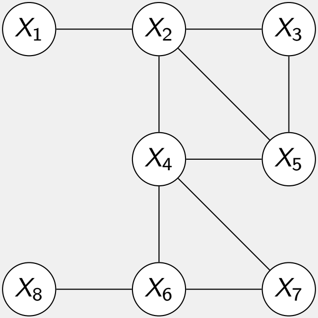


Cluster graph

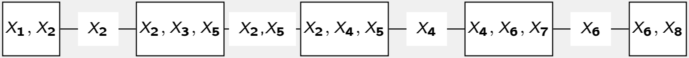


Belief propagation: message passing

Markov network



Clique tree



Cluster graph

Alternative representation of the probability distribution

Cluster graph for a set of variables, $\boldsymbol{X}$

A *cluster graph* $\mathcal{U} = (V_{\mathcal{U}}, E_{\mathcal{U}})$ over $\boldsymbol{X}$ is an undirected graph s.t.:

- Each node i is associated with a subset $\boldsymbol{C}_i \subset \boldsymbol{X}$ (cluster)
- Each edge i, j between clusters $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ is associated with a sepset $\boldsymbol{S}_{ij} \subseteq \boldsymbol{C}_i \cap \boldsymbol{C}_j$
- A set of beliefs is considered:
  - There is a factor β_i over $\boldsymbol{C}_i$ for each cluster $\boldsymbol{C}_i$
  - There is a factor μ_{ij} over $\boldsymbol{S}_{ij}$ for each sepset $\boldsymbol{S}_{ij}$
- The encoded probability distribution is:

$$P_{\mathcal{U}}(X) \propto \frac{\prod_{i \in V_{\mathcal{U}}} \beta_i(\boldsymbol{C}_i)}{\prod_{\{i,j\} \in E_{\mathcal{U}}} \mu_{ij}(\boldsymbol{S}_{i,j})} = \prod_{i=1}^{f} \phi_f$$

Cluster graph

A few examples

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z)}{\mu_{12}(Y)}$$

Cluster graph

A few examples

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z)}{\mu_{12}(Y)}$$

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z) \times \beta_3(Z, X)}{\mu_{12}(Y) \times \mu_{13}(X) \times \mu_{23}(Z)}$$

Cluster graph

A few examples

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z)}{\mu_{12}(Y)}$$

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z) \times \beta_3(Z, X)}{\mu_{12}(Y) \times \mu_{13}(X) \times \mu_{23}(Z)}$$

$$P(X, Y, Z, A) \propto \frac{\beta_1(X, Y, Z) \times \beta_2(Y, Z, A)}{\mu_{12}(Y, Z)}$$

Cluster graph

A few examples

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z)}{\mu_{12}(Y)}$$

$$P(X, Y, Z) \propto \frac{\beta_1(X, Y) \times \beta_2(Y, Z) \times \beta_3(Z, X)}{\mu_{12}(Y) \times \mu_{13}(X) \times \mu_{23}(Z)}$$

$$P(X, Y, Z, A) \propto \frac{\beta_1(X, Y, Z) \times \beta_2(Y, Z, A)}{\mu_{12}(Y, Z)}$$

$$P(X, Y, Z, A, B) \propto \frac{\beta_1(X, Y, Z) \times \beta_2(Y, Z, A) \times \beta_3(A, B, X)}{\mu_{12}(Y, Z) \times \mu_{23}(A)}$$

Cluster graph

Alternative representation of the probability distribution

Result:

We build cluster factors β_i such that:

$$p(X_j) = \frac{1}{Z} \sum_{\boldsymbol{X} \setminus X_j} \prod_{l=1}^{f} \phi_l$$

$$= \frac{1}{Z} \sum_{\boldsymbol{C_i} \setminus X_j} \beta_i(\boldsymbol{C}_i)$$

such that $X_j \in \boldsymbol{C}_i$ (and usually $|\boldsymbol{C}_i| << |\boldsymbol{X}|$).

Cluster graph

Alternative representation of the probability distribution

Result:

We build cluster factors β_i such that:

$$p(X_j) = \frac{1}{Z} \sum_{\boldsymbol{X} \setminus X_j} \prod_{l=1}^{f} \phi_l$$

$$= \frac{1}{Z} \sum_{\boldsymbol{C_i} \setminus X_j} \beta_i(\boldsymbol{C}_i)$$

such that $X_j \in \boldsymbol{C}_i$ (and usually $|\boldsymbol{C}_i| << |\boldsymbol{X}|$).

So, how we build these cluster factors, β_i?

Cluster graph

Family preservation property

Let $P(\boldsymbol{X})$ be a distribution that factorizes as follow,

$$P(\boldsymbol{X}) = \frac{1}{\Theta} \prod_{\phi \in \Phi} \phi$$

To represent a probability distribution $P(\boldsymbol{X})$ by means of a cluster graph $\mathcal{U}$, **family preservation** is required:

Each **factor** $\phi \in \Phi$ is associated with a **cluster** C_i such that $Scope(\phi) \subseteq C_i$

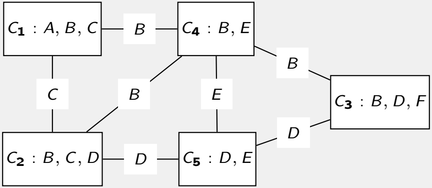
Cluster graph

Family preservation property

Cluster graph for $\boldsymbol{X} = \{A, B, C, D, E, F\}$

$$P(\boldsymbol{X}) \propto$$

$$\phi_1(A, B, C)\phi_2(B, C)\phi_3(B, D)\phi_4(D, E)\phi_5(B, E)\phi_6(B, D, F)$$



Cluster graph

Running intersection property (RIP)

A cluster graph $\mathcal{U}$ satisfies the **running intersection property** if,

for any variable X such that $X \in \boldsymbol{C}_i$ and $X \in \boldsymbol{C}_j$ ($\boldsymbol{C}_i \neq \boldsymbol{C}_j$),

there exists a unique path between $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ $(\boldsymbol{C}_i, \boldsymbol{C}_k, \boldsymbol{C}_{k+1}, \dots, \boldsymbol{C}_{k+l}, \boldsymbol{C}_j)$

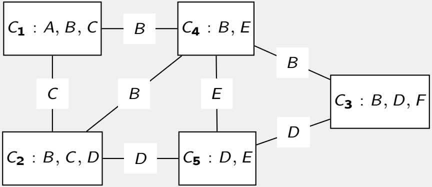
such that X is in every intermediate cluster and sepset:

- $X \in \boldsymbol{C}_k$, for all $\{k, k+1, \dots\}$
- $X \in \boldsymbol{S}_{k,k+1}$, for all $\{k, k+1, \dots\}$ and $X \in \boldsymbol{S}_{ik} \wedge X \in \boldsymbol{S}_{k+l,j}$

Cluster graph

Running intersection property (RIP)

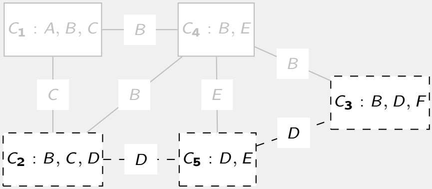
Cluster graph for $\boldsymbol{X} = \{A, B, C, D, E, F\}$



Cluster graph

Running intersection property (RIP)

Cluster graph for $\boldsymbol{X} = \{A, B, C, D, E, F\}$



E.g., $X = D$

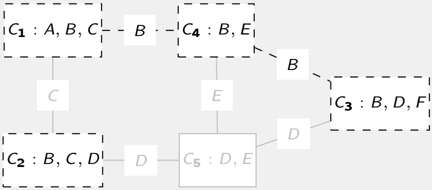
Graphical viewpoint:

If we build a subgraph by removing all the clusters and sepsets that do not contain X, the remaining subgraph is connected and has no loop

Cluster graph

Running intersection property (RIP)

Cluster graph for $\boldsymbol{X} = \{A, B, C, D, E, F\}$

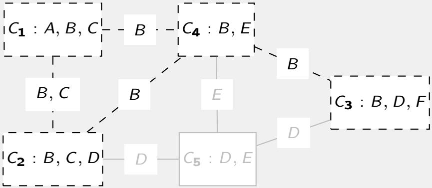


No path for B from $\boldsymbol{C}_2$ in this new cluster graph

Cluster graph

Running intersection property (RIP)

Cluster graph for $\boldsymbol{X} = \{A, B, C, D, E, F\}$

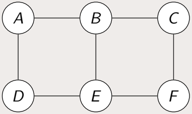


Loop for B in $\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_4$ (path not unique) in this new cluster graph

Exercise

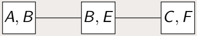
Cluster Graph construction

Given the following MN,

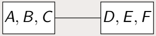


which is a valid cluster graph for it?

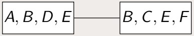
a)



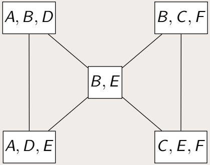
b)



c)



d)



Exercise

Family Preservation

Suppose we have a factor $P(A|C)$ that we wish to include in our sum-product message passing inference. We should:

a) Assign the factor to all cliques that contain A or C
b) Assign the factor to all cliques that contain A and C
c) Assign the factor to one clique that contain A and C
d) None of these

Clique tree

Definition

A **clique tree**, $\mathcal{T}$, is a cluster tree that satisfies the *running intersection property.*
A **cluster tree** is a cluster graph without loops. It satisfies the RIP if this equality always holds

$$\boldsymbol{S}_{ij} = \boldsymbol{C}_i \cap \boldsymbol{C}_j$$

Independence

$$\boldsymbol{W}_i = \bigcup_{k \text{ has path to } i \text{ and } k \neq j} \boldsymbol{C}_k \qquad \boldsymbol{W}_j = \bigcup_{k \text{ has path to } j \text{ and } k \neq i} \boldsymbol{C}_k$$

Then,

$$\{\boldsymbol{W}_i \backslash \boldsymbol{S}_{ij}\} \perp\!\!\!\perp \{\boldsymbol{W}_j \backslash \boldsymbol{S}_{ij}\} | \boldsymbol{S}_{ij}$$

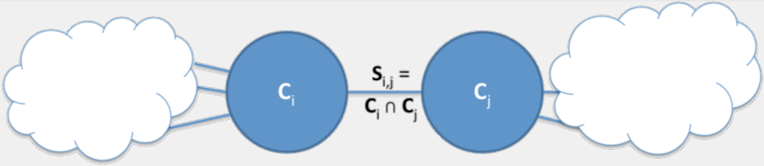
Clique tree

Definition

A **clique tree**, $\mathcal{T}$, is a cluster tree that satisfies the *running intersection property.*
A **cluster tree** is a cluster graph without loops. It satisfies the RIP if this equality always holds

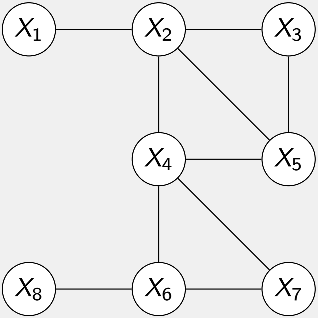
$$\boldsymbol{S}_{ij} = \boldsymbol{C}_i \cap \boldsymbol{C}_j$$

Independence

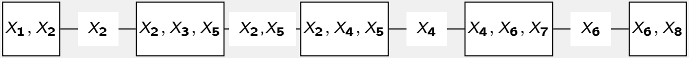


Belief propagation: message passing

Markov network

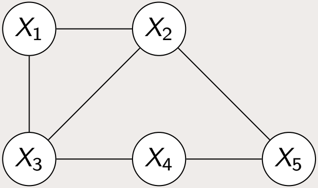


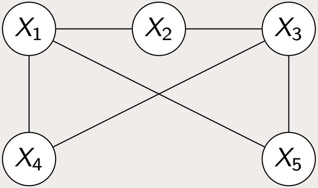
Clique tree



Exercise

Draw the clique tree





Exact Inference

Probabilistic Graphical Models

Jerónimo Hernández-González

Belief propagation: message passing

Sum-Product algorithm over Clique trees

Whereas in VE variables are removed one by one, BP provides a more general way to assess marginals

It can be easily extended to multiple simultaneous queries

Belief propagation: message passing

Sum-Product algorithm over Clique trees

Cluster factor

$$\psi_i(\boldsymbol{C}_i) = \prod_{\alpha(\phi)=\boldsymbol{C}_i} \phi$$

where $\alpha(\phi)$ is a function that assigns each factor ϕ in the original model $\mathcal{H}$ to a cluster C_i in $\mathcal{T}$

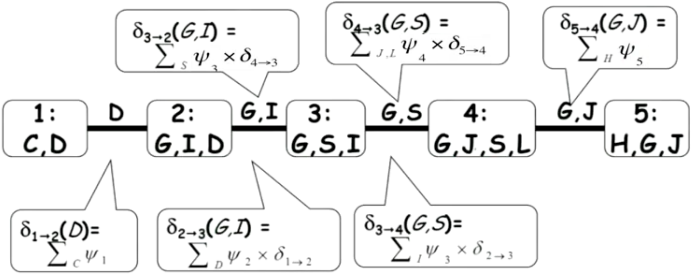
Message

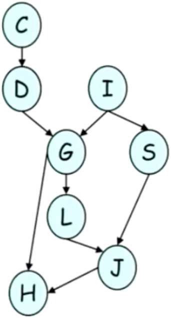
$$\delta_{i\to j}(\boldsymbol{S}_{ij}) = \sum_{\boldsymbol{C}_i \setminus \boldsymbol{S}_{ij}} \left(\psi_i \prod_{k \in \{\boldsymbol{d}_i - j\}} \delta_{k\to i} \right)$$

where $\boldsymbol{d}_i$ is the set of clusters directly connected to $\boldsymbol{C}_i$

Belief propagation: message passing

Sum-Product algorithm over Clique trees





Belief propagation: message passing

Sum-Product algorithm over Clique trees

1. Select a root clique C_r
2. Starting from the leaves and up to the root, each node pass the messages:

$$\delta_{i \to j} = \sum_{\boldsymbol{C}_i \setminus \boldsymbol{S}_{ij}} \left(\psi_i \prod_{k \in \{\boldsymbol{d}_i - j\}} \delta_{k \to i} \right)$$

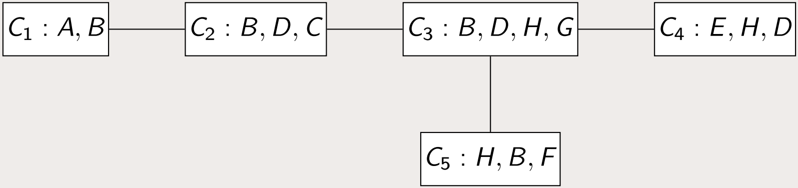
3. Select another root and repeat the process until messages are passed in both direction throughout all edges
4. The cluster and sepset beliefs are assessed, resp., as

$$\beta_i(\boldsymbol{C}_i) = \psi_i \prod_{j \in \boldsymbol{d}_i} \delta_{j \to i} \qquad \mu_{ij}(\boldsymbol{S}_{ij}) = \delta_{i \to j} \delta_{j \to i}$$

Exercise

Message Ordering

In this clique tree



which of the following starting message passing orders is/are valid?

a) $C_1 \rightarrow C_2,\ C_2 \rightarrow C_3,\ C_3 \rightarrow C_4,\ C_3 \rightarrow C_5$

b) $C_4 \rightarrow C_3,\ C_3 \rightarrow C_2,\ C_2 \rightarrow C_1$

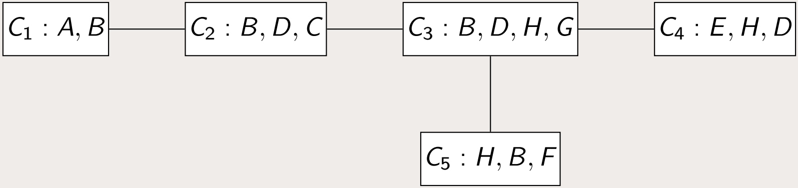
c) $C_4 \rightarrow C_3,\ C_5 \rightarrow C_3,\ C_2 \rightarrow C_3$

d) $C_1 \rightarrow C_2,\ C_2 \rightarrow C_3,\ C_5 \rightarrow C_3,\ C_3 \rightarrow C_4$

Exercise

Message Passing in a Clique Tree

In this clique tree



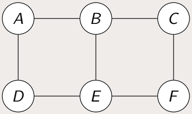
Which is the correct form of the message from clique 3 to clique 2, $\delta_{3\to2}$, where $\psi_i(C_i)$ is the initial potential of clique i?

a) $\sum_{G,H} \psi_3(C_3) \times \delta_{4\to3} \times \delta_{5\to3}$
b) $\sum_{B,D} \psi_3(C_3) \times \delta_{4\to3} \times \delta_{5\to3}$
c) $\sum_{B,D,G,H} \psi_3(C_3) \times \delta_{4\to3} \times \delta_{5\to3}$
d) $\sum_{G,H} \psi_3(C_3) \times \delta_{2\to3}$

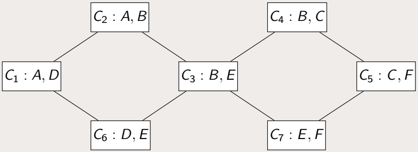
Exercise

Message Passing in a Cluster Graph

To perform inference in this MN, we use this Cluster Graph:



MN



CG

Which expression correctly represents the message $\delta_{3\to6}$?

a) $\delta_{3\to6}(E) = \sum_B \phi_{B,E}(B, E) \cdot \delta_{2\to3}(B) \cdot \delta_{4\to3}(B) \cdot \delta_{7\to3}(E) \cdot \delta_{6\to3}(B)$

b) $\delta_{3\to6}(E) = \sum_B \phi_{B,E}(B, E) \cdot \delta_{2\to3}(B) \cdot \delta_{4\to3}(B) \cdot \delta_{7\to3}(E)$

c) $\delta_{3\to6}(B, E) = \phi_{B,E}(B, E) \cdot \delta_{2\to3}(B) \cdot \delta_{4\to3}(B) \cdot \delta_{7\to3}(E)$

d) $\delta_{3\to6}(E) = \sum_B \delta_{2\to3}(B) \cdot \delta_{4\to3}(B) \cdot \delta_{7\to3}(E)$

* Assume that the vars. in the sepsets are equal to the intersection of the vars. in the linked cliques

Belief propagation: message passing

Sum-Product algorithm over Clique trees

Cluster graph (rev.)

- The encoded probability distribution is:

$$P_{\mathcal{U}}(X) \propto \frac{\prod_{i \in V_{\mathcal{U}}} \beta_i(\boldsymbol{C}_i)}{\prod_{\{i,j\} \in E_{\mathcal{U}}} \mu_{ij}(\boldsymbol{S}_{ij})}$$

- Expanded as

$$\begin{aligned} P_{\mathcal{U}}(X) &\propto \frac{\prod_{i \in V_{\mathcal{U}}} \psi_i \prod_{j \in d_i} \delta_{j \to i}}{\prod_{\{i,j\} \in E_{\mathcal{U}}} \delta_{i \to j} \delta_{j \to i}} \\ &= \prod_{i \in V_{\mathcal{U}}} \psi_i = \prod_{\boldsymbol{C}_i} \prod_{\alpha(\phi) = \boldsymbol{C}_i} \phi \\ &= \prod_{i=1}^{f} \phi_i \end{aligned}$$

Belief propagation: message passing

Sum-Product algorithm over Clique trees

Calibration

An egde i, j in a clique tree $\mathcal{T}$ is calibrated when:

$$\sum_{\mathbf{C}_i \setminus \mathbf{S}_{i,j}} \beta_i = \sum_{\mathbf{C}_j \setminus \mathbf{S}_{i,j}} \beta_j$$

A clique tree $\mathcal{T}$ is calibrated when all its edges are calibrated

Queries in a calibrated clique tree:

- the marginal $p(X)$ is simply the marginalized and normalized belief β_i of a clique that contains X ($X \in \mathbf{C}_i$): $p(X) = Norm[\sum_{\mathbf{C}_i \setminus X} \beta_i(\mathbf{C}_i)]$
- the conditional query $p(X|E = e)$ can also be assessed efficiently (two cases: same clique or not)

Belief propagation: message passing

Sum-Product algorithm over Clique trees

Clique Tree algorithm, up to our knowledge,

- Can solve the same queries as variable elimination.
- Only takes a small advantage of the opportunity to remove several variables at once.

Then, why did we get into this clique tree business?

Belief propagation: message passing

Sum-Product algorithm over Clique trees

Many simultaneous CPQ

Given a distribution $P(\boldsymbol{X})$ that factorizes as

$$P(\boldsymbol{X}) = \frac{1}{\Theta} \prod_{i=1}^{f} \phi_i.$$

a set of queries $\{\boldsymbol{Y}_1, \dots \boldsymbol{Y}_k\}$ and $\boldsymbol{Z}_j = \boldsymbol{X} \setminus \boldsymbol{Y}_j$, assess

$$P(\boldsymbol{Y}_j) = \sum_{\boldsymbol{Z}_j} P(\boldsymbol{Y}_j, \boldsymbol{Z}_j) = \sum_{\boldsymbol{Z}_j} \prod_{i=1}^{f} \phi_i \qquad \forall j \in \{1, \dots, k\}$$

Calibrated clique trees are very efficient to assess many different marginals of the same distribution

Summary

Inference, or answering a conditional probability query

Exact inference: Only with induced graphs of reduced width

- VE answers a conditional probability query in polynomial time
- If we have more than one CPQ over the same distribution:
  1. Represent the distribution as a clique tree
  2. Calibrate the tree
  3. Use it to efficiently answer the queries
- In the **general case**, inference is not polynomial

Belief propagation: message passing

General case

Loopy belief propagation in cluster graphs

- Assign each factor ϕ_i to a cluster $\boldsymbol{C}_{\alpha(\phi_i)}$
- Construct factors $\psi_i(\boldsymbol{C}_i) = \prod_{\phi:\alpha(\phi)=\boldsymbol{C}_i} \phi$
- Initialize all messages to 1
- Repeat
  - Select an edge i, j
  - Pass message $\delta_{i\to j}(\boldsymbol{S}_{ij}) = \sum_{\boldsymbol{C}_i \setminus \boldsymbol{S}_{ij}} \left(\psi_i \prod_{k\in\{\boldsymbol{d}_i - j\}} \delta_{k\to i}\right)$
- Obtain beliefs $\beta_i(\boldsymbol{C}_i) = \psi_i(\boldsymbol{C}_i) \prod_{k\in\boldsymbol{d}_i} \delta_{k\to i}$

** Approximate in most cases

Exact Inference

Probabilistic Graphical Models

Jerónimo Hernández-González