Exact Inference

Variable Elimination

Wilker Aziz w.aziz@uva.nl

Fall 2025

https://probabll.github.io

Outline and goals

Module 3 introduces *a key algorithm for Exact Inference* known as **Variable Elimination** (VE; Chapter 9).

ILOs After this module the student

- can perform sum-product inference via VE;
- can predict the complexity of VE by constructing an induced graph;
- can optimise VE by ordering elimination steps opportunistically.

Textbook for this course: Koller and Friedman [1].

Overview of Module 3

HC4a: naive sum-product inference and variable elimination.

LC4: VE in code.

HC4b: elimination ordering.

WC4: exercises.

Table of contents

Naive Exact Inference

We have a joint distribution $P(\boldsymbol{X})$ over a collection of rvs $\boldsymbol{X}$, this distribution is represented by a PGM such as a BN or MN.

Conditional probability queries:

- **Evidence:** an assignment $\boldsymbol{E} = \boldsymbol{e}$ of a subset $\boldsymbol{E} \subseteq \boldsymbol{X}$ of the rvs.
- **Query:** given the evidence, we are interest in reasoning about the possible assignments of the rvs in $\boldsymbol{Q} \subseteq \boldsymbol{X}$, where $\boldsymbol{Q} \cap \boldsymbol{E} = \emptyset$.
- **Task:** express $P(\boldsymbol{Q}|\boldsymbol{E} = \boldsymbol{e})$

Applications: medical/fault diagnosis, image/text completion, speech analysis, . . .

$$P(\mathbf{Q}|\mathbf{E}=\mathbf{e}) = \frac{P(\mathbf{Q},\mathbf{E}=\mathbf{e})}{P(\mathbf{E}=\mathbf{e})} \quad \propto P(\mathbf{Q},\mathbf{E}=\mathbf{e})$$

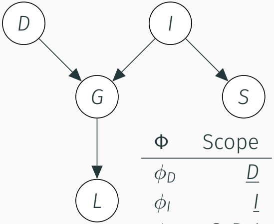
Let $\mathbf{W}$ be the complement of $\mathbf{E}$ and $\mathbf{Q}$ in $\mathbf{X}$. That is, the 'rest' of the variables in the model.

Then, we can express $P(\mathbf{Q},\mathbf{E}=\mathbf{e})$ via marginalisation of $\mathbf{W}$. This is an instance of what we call inference. Specifically, *marginal inference*:

$$P(\mathbf{Q},\mathbf{E}=\mathbf{e}) = \sum_{\mathbf{W}} P(\mathbf{W},\mathbf{Q},\mathbf{E}=\mathbf{e})$$

To continue, we need to know more about how $P(\underbrace{\mathbf{W},\mathbf{Q},\mathbf{E}}_{\mathbf{X}})$ factorises.

A factor-view of the *Student* example



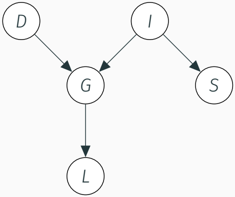
Φ	Scope
ϕ_D	$\underline{D}$
ϕ_I	$\underline{I}$
ϕ_G	$\underline{G}, D, I$
ϕ_S	$\underline{S}, I$
ϕ_L	$\underline{L}, G$

In factor form, we use ϕ_X with scope $\{X\} \cup \mathrm{Pa}(X)$ for $P(X|\mathrm{Pa}(X))$.

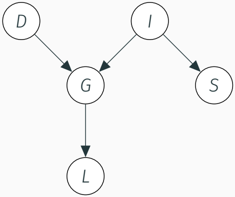
Joint distribution:

$$\begin{aligned} P(D, I, S, G, L) &= P(D)P(I)P(S|I)P(G|D, I)P(L|G) \\ &= \phi_D(D)\phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G) \end{aligned}$$

In this view, we can develop a framework for inference that works both for BNs and MNs.



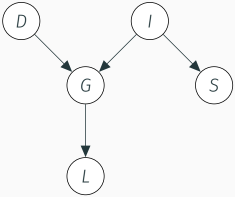
Evidence: $D = d^1$. Query: L.



Evidence: $D = d^1$. Query: L.

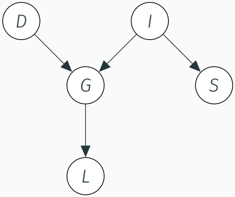
To express $P(L|D = d^1)$

1. assign $D = d^1$:
   $P(D = d^1, I, S, G, L)$



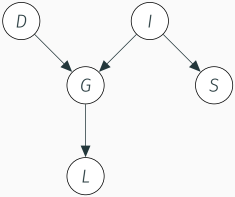
Evidence: $D = d^1$. Query: L.
To express $P(L|D = d^1)$

1. assign $D = d^1$:
   $P(D = d^1, I, S, G, L)$
2. marginalise I, G, S:
   $P(L, D = d^1)$;



Evidence: $D = d^1$. Query: L.
To express $P(L|D = d^1)$

1. assign $D = d^1$:
   $P(D = d^1, I, S, G, L)$
2. marginalise I, G, S:
   $P(L, D = d^1)$;
3. normalize over $\text{Val}(L)$.

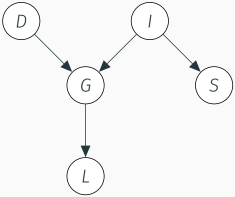


Evidence: $D = d^1$. Query: L.
To express $P(L|D = d^1)$

1. assign $D = d^1$:
   $P(D = d^1, I, S, G, L)$
2. marginalise I, G, S:
   $P(L, D = d^1)$;
3. normalize over $\mathrm{Val}(L)$.

Start with the joint distribution, but let's write it using *factors*:

$$P(D, I, G, S, L)$$
$$= \phi_D(D)\phi_I(I)\phi_G(G, D, I)\phi_S(S, I)\phi_L(L, G)$$



Evidence: $D = d^1$. Query: L.
To express $P(L|D = d^1)$

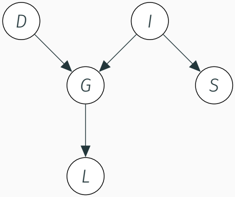
1. assign $D = d^1$:
   $P(D = d^1, I, S, G, L)$
2. marginalise I, G, S:
   $P(L, D = d^1)$;
3. normalize over $\mathrm{Val}(L)$.

Start with the joint distribution, but let's write it using *factors*:

$$P(D, I, G, S, L)$$
$$= \phi_D(D)\phi_I(I)\phi_G(G, D, I)\phi_S(S, I)\phi_L(L, G)$$

1. Reduce the factors to the available evidence:

$$P(D = d^1, I, G, S, L)$$
$$= \phi_D[d^1]\phi_I(I)\phi_G[d^1](G, I)\phi_S(S, I)\phi_L(L, G)$$



Evidence: $D = d^1$. Query: L.
To express $P(L|D = d^1)$

1. assign $D = d^1$:
   $P(D = d^1, I, S, G, L)$
2. marginalise I, G, S:
   $P(L, D = d^1)$;
3. normalize over $\mathrm{Val}(L)$.

Start with the joint distribution, but let's write it using *factors*:

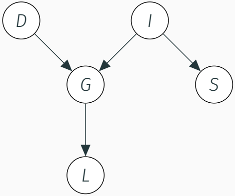
$$P(D, I, G, S, L) = \phi_D(D)\phi_I(I)\phi_G(G, D, I)\phi_S(S, I)\phi_L(L, G)$$

1. Reduce the factors to the available evidence:

$$P(D = d^1, I, G, S, L) = \phi_D[d^1]\phi_I(I)\phi_G[d^1](G, I)\phi_S(S, I)\phi_L(L, G)$$

2. Marginalise the variables that are neither assigned nor queried about:

$$P(L, D = d^1) = \sum_{G,I,S} \phi_D[d^1]\phi_I(I)\phi_G[d^1](G, I)\phi_S(S, I)\phi_L(L, G)$$



Evidence: $D = d^1$. Query: L.
To express $P(L|D = d^1)$

1. assign $D = d^1$: $P(D = d^1, I, S, G, L)$
2. marginalise I, G, S: $P(L, D = d^1)$;
3. normalize over $\mathrm{Val}(L)$.

Start with the joint distribution, but let's write it using *factors*:

$$P(D, I, G, S, L) \\ = \phi_D(D)\phi_I(I)\phi_G(G, D, I)\phi_S(S, I)\phi_L(L, G)$$

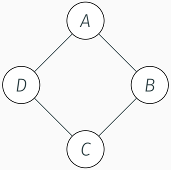
1. Reduce the factors to the available evidence:

$$P(D = d^1, I, G, S, L) \\ = \phi_D[d^1]\phi_I(I)\phi_G[d^1](G, I)\phi_S(S, I)\phi_L(L, G)$$

2. Marginalise the variables that are neither assigned nor queried about:

$$P(L, D = d^1) \\ = \sum_{G,I,S} \phi_D[d^1]\phi_I(I)\phi_G[d^1](G, I)\phi_S(S, I)\phi_L(L, G)$$

3. Normaliser: $\sum_{l \in \mathrm{Val}(L)} P(L = l, D = d^1)$



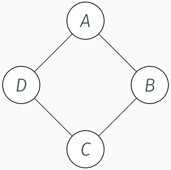
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Evidence: $A = a^1, C = c^1$.

Query: B.

To express $P_\Phi(B|A = a^1, C = c^1)$

1. assign $A = a^1, C = c^1$:
   $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$



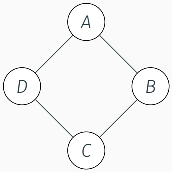
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Evidence: $A = a^1, C = c^1$.

Query: B.

To express $P_\Phi(B|A = a^1, C = c^1)$

1. assign $A = a^1, C = c^1$:
   $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$
2. marginalise D:
   $\tilde{P}_\Phi(B, A = a^1, C = c^1)$;



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

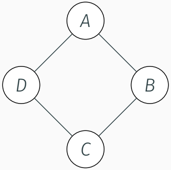
Evidence: $A = a^1, C = c^1$.

Query: B.

To express $P_\Phi(B|A = a^1, C = c^1)$

1. assign $A = a^1, C = c^1$:
   $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$
2. marginalise D:
   $\tilde{P}_\Phi(B, A = a^1, C = c^1)$;
3. normalize over $\mathrm{Val}(B)$.

Exact Inference: Sum-Product — *Misconception* example



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Evidence: $A = a^1, C = c^1$.

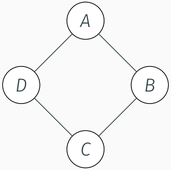
Query: B.

To express $P_\Phi(B|A = a^1, C = c^1)$

1. assign $A = a^1, C = c^1$:
   $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$
2. marginalise D:
   $\tilde{P}_\Phi(B, A = a^1, C = c^1)$;
3. normalize over $\mathrm{Val}(B)$.

Start with the unnormalised measure:

$$\tilde{P}_\Phi(A, B, C, D)$$
$$= \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Evidence: $A = a^1, C = c^1$.
Query: B.
To express $P_\Phi(B|A = a^1, C = c^1)$

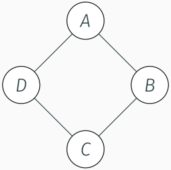
1. assign $A = a^1, C = c^1$:
   $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$
2. marginalise D:
   $\tilde{P}_\Phi(B, A = a^1, C = c^1)$;
3. normalize over $\mathrm{Val}(B)$.

Start with the unnormalised measure:

$$\tilde{P}_\Phi(A, B, C, D)$$
$$= \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

1. Reduce factors to the evidence:

$$\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$$
$$= \phi_1[a^1](B)\phi_2[c^1](B)\phi_3[c^1](D)\phi_4[a^1](D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Evidence: $A = a^1, C = c^1$.
Query: B.
To express $P_\Phi(B|A = a^1, C = c^1)$

1. assign $A = a^1, C = c^1$: $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$
2. marginalise D: $\tilde{P}_\Phi(B, A = a^1, C = c^1)$;
3. normalize over $\mathrm{Val}(B)$.

Start with the unnormalised measure:

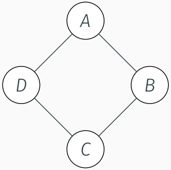
$$\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

1. Reduce factors to the evidence:

$$\tilde{P}_\Phi(A = a^1, B, C = c^1, D) = \phi_1[a^1](B)\phi_2[c^1](B)\phi_3[c^1](D)\phi_4[a^1](D)$$

2. Marginalise the variables that are neither assigned nor queried about:

$$\tilde{P}_\Phi(B, A = a^1, C = c^1) = \sum_D \phi_1[a^1](B)\phi_2[c^1](B)\phi_3[c^1](D)\phi_4[a^1](D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Evidence: $A = a^1, C = c^1$.
Query: B.
To express $P_\Phi(B|A = a^1, C = c^1)$

1. assign $A = a^1, C = c^1$:
   $\tilde{P}_\Phi(A = a^1, B, C = c^1, D)$
2. marginalise D:
   $\tilde{P}_\Phi(B, A = a^1, C = c^1)$;
3. normalize over $\mathrm{Val}(B)$.

Start with the unnormalised measure:

$$\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

1. Reduce factors to the evidence:

$$\tilde{P}_\Phi(A = a^1, B, C = c^1, D) = \phi_1[a^1](B)\phi_2[c^1](B)\phi_3[c^1](D)\phi_4[a^1](D)$$

2. Marginalise the variables that are neither assigned nor queried about:

$$\tilde{P}_\Phi(B, A = a^1, C = c^1) = \sum_D \phi_1[a^1](B)\phi_2[c^1](B)\phi_3[c^1](D)\phi_4[a^1](D)$$

3. Normaliser:
$\sum_{b \in \mathrm{Val}(B)} \tilde{P}_\Phi(B = b, A = a^1, C = c^1)$

Exact Inference

Naive marginal inference requires creating the full joint (unnormalised) distribution—a large product—before we can marginalise (sum). Hence the name 'sum-product'.

Of course, we can reduce factors to the evidence before taking the large product, but you can see that this is not an efficient algorithm.

No matter the graphical structure, it will always involve a product of all factors, which, in the worst case, is a very large factor.

Next, we use the graphical structure to obtain better average performance via the so-called variable elimination algorithm.

Variable Elimination

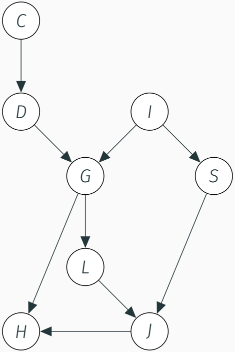
Exploiting the Graph Structure

Naive sum-product does not use the graph structure at all. It starts from a full product of all factors.

The key challenge in inference is marginalisation of unassigned rvs. We build the entire product of factors because then we are sure that it is safe to marginalise rvs.

But consider marginalising one rv at a time, by looking around in the graphical structure, we might be able to see that not all factors are going to be relevant.

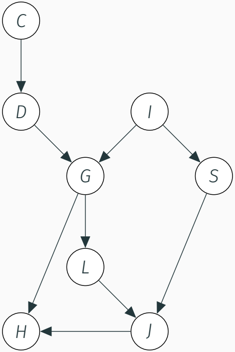
This intuition is the key to a more 'incremental' procedure known as variable elimination (VE; Section 9.3).



In factor form, we use ϕ_X with scope $\{X\} \cup \mathrm{Pa}(X)$ for $P(X|\mathrm{Pa}(X))$.

Joint distribution:

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

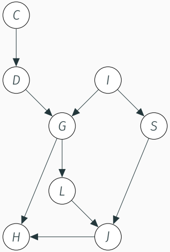


In factor form, we use ϕ_X with scope $\{X\} \cup \mathrm{Pa}(X)$ for $P(X|\mathrm{Pa}(X))$.

Joint distribution:

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

Example: let's express $P(J)$ by eliminating C, D, I, H, G, S, L in this order [example 9.1]



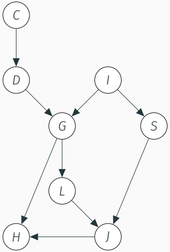
Eliminate: $\underline{C}, D, I, H, G, S, L$.

Joint distribution:

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

Goal: $P(J)$.

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



Eliminate: $\underline{C}, D, I, H, G, S, L$.

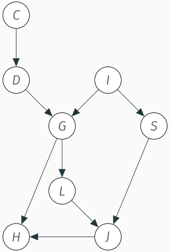
Joint distribution:

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

Goal: $P(J)$.

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

$$= \phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J) \underbrace{\sum_C \overbrace{\phi_C(C)\phi_D(D, C)}^{\rho_1(C,D)}}_{\tau_1(D)}$$



Eliminate: $\underline{C}, D, I, H, G, S, L$.

Joint distribution:

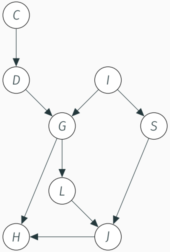
$$P(C, D, I, G, S, L, JH) = \phi_C(C)\phi_D(D, C)\phi_I(I)\phi_S(S, I) \times \phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)$$

Goal: $P(J)$.

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

$$= \phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J) \underbrace{\sum_C \overbrace{\phi_C(C)\phi_D(D, C)}^{\rho_1(C,D)}}_{\tau_1(D)}$$

$$= \phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_1(D)$$



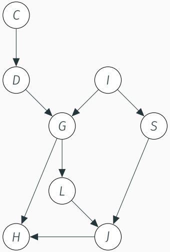
Eliminate: $\cancel{C}, \underline{D}, I, H, G, S, L$.

Joint distribution:

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

Goal: $P(J)$.

$$\sum_D \phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_1(D)$$



Joint distribution:

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

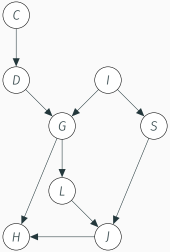
Goal: $P(J)$.

Eliminate: $\cancel{C}$, $\underline{D}$, I, H, G, S, L.

$$\sum_D \phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_1(D)$$

$$= \phi_I(I)\phi_S(S, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\underbrace{\sum_D \overbrace{\tau_1(D)\phi_G(G, D, I)}^{\rho_2(G,D,I)}}_{\tau_2(G,I)}$$

$P(J)$ for *Extended Student* — Eliminate D



Joint distribution:

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

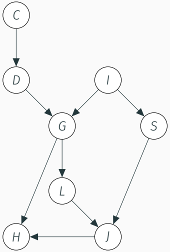
Goal: $P(J)$.

Eliminate: $\cancel{C}, \underline{D}, I, H, G, S, L$.

$$\sum_D \phi_I(I)\phi_S(S, I)\phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_1(D)$$

$$= \phi_I(I)\phi_S(S, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J) \underbrace{\sum_D \overbrace{\tau_1(D)\phi_G(G, D, I)}^{\rho_2(G,D,I)}}_{\tau_2(G,I)}$$

$$= \phi_I(I)\phi_S(S, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_2(G, I)$$



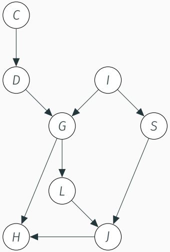
Eliminate: $\cancel{C}, \cancel{D}, \underline{I}, H, G, S, L$.

Joint distribution:

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

Goal: $P(J)$.

$$\sum_I \phi_I(I)\phi_S(S, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_2(G, I)$$



Eliminate: $\cancel{C}, \cancel{D}, \underline{I}, H, G, S, L$.

Joint distribution:

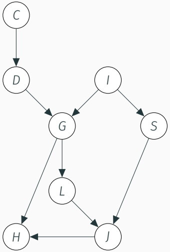
$$P(C, D, I, G, S, L, JH) = \phi_C(C)\phi_D(D, C)\phi_I(I)\phi_S(S, I) \times \phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)$$

Goal: $P(J)$.

$$\sum_I \phi_I(I)\phi_S(S, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_2(G, I)$$

$$= \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J) \underbrace{\sum_I \overbrace{\phi_I(I)\phi_S(S, I)\tau_2(G, I)}^{\rho_3(I,G,S)}}_{\tau_3(G,S)}$$

$P(J)$ for *Extended Student* — Eliminate I



Eliminate: $\cancel{C}, \cancel{D}, \underline{I}, H, G, S, L$.

Joint distribution:

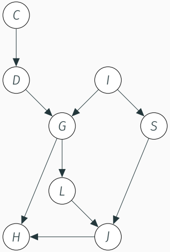
$$P(C, D, I, G, S, L, JH) = \phi_C(C)\phi_D(D, C)\phi_I(I)\phi_S(S, I) \times \phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)$$

Goal: $P(J)$.

$$\sum_I \phi_I(I)\phi_S(S, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_2(G, I)$$

$$= \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J) \underbrace{\sum_I \overbrace{\phi_I(I)\phi_S(S, I)\tau_2(G, I)}^{\rho_3(I,G,S)}}_{\tau_3(G,S)}$$

$$= \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_3(G, S)$$



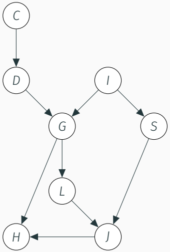
Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \underline{H}, G, S, L$.

Joint distribution:

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

Goal: $P(J)$.

$$\sum_H \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_3(G, S)$$



Joint distribution:

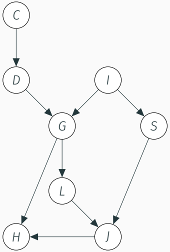
$$P(C, D, I, G, S, L, JH) = \phi_C(C)\phi_D(D, C)\phi_I(I)\phi_S(S, I) \\ \times \phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)$$

Goal: $P(J)$.

Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \underline{H}, G, S, L$.

$$\sum_H \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_3(G, S)$$
$$= \phi_L(L, G)\phi_J(J, L)\tau_3(G, S) \underbrace{\sum_H \phi_H(H, G, J)}_{\tau_4(G,J)}$$

$P(J)$ for *Extended Student* — Eliminate H



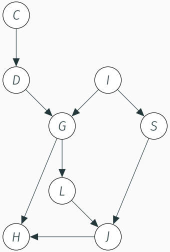
Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \underline{H}, G, S, L$.

Joint distribution:

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

Goal: $P(J)$.

$$\begin{aligned}
&\sum_H \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_3(G, S) \\
&= \phi_L(L, G)\phi_J(J, L)\tau_3(G, S) \underbrace{\sum_H \phi_H(H, G, J)}_{\tau_4(G,J)} \\
&= \phi_L(L, G)\phi_J(J, L)\tau_3(G, S)\tau_4(G, J)
\end{aligned}$$



Joint distribution:

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

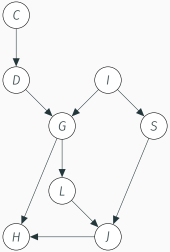
Goal: $P(J)$.

Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \underline{H}, G, S, L$.

$$\begin{aligned} &\sum_H \phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)\tau_3(G, S) \\ &= \phi_L(L, G)\phi_J(J, L)\tau_3(G, S) \underbrace{\sum_H \phi_H(H, G, J)}_{\tau_4(G,J)} \\ &= \phi_L(L, G)\phi_J(J, L)\tau_3(G, S)\tau_4(G, J) \end{aligned}$$

$\tau_4(G, J) = 1$ but, for generality, let's pretend we do not know that.

$P(J)$ for *Extended Student* — Eliminate G



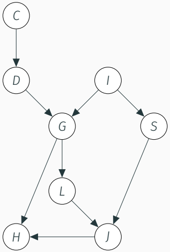
Joint distribution:

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

Goal: $P(J)$.

Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \underline{G}, S, L$.

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



Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \underline{G}, S, L$.

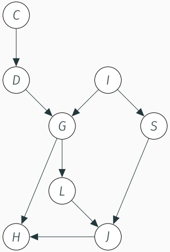
Joint distribution:

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

Goal: $P(J)$.

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

$$= \underbrace{\phi_J(J, L)\sum_G \overbrace{\phi_L(L, G)\tau_3(G, S)\tau_4(G, J)}^{\rho_5(G,J,L,S)}}_{\tau_5(J,L,S)}$$



Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \underline{G}, S, L$.

Joint distribution:

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

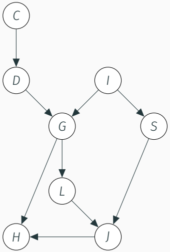
Goal: $P(J)$.

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

$$= \phi_J(J, L) \underbrace{\sum_G \overbrace{\phi_L(L, G)\tau_3(G, S)\tau_4(G, J)}^{\rho_5(G,J,L,S)}}_{\tau_5(J,L,S)}$$

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

$P(J)$ for *Extended Student* — Eliminate S



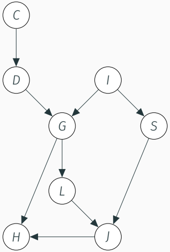
Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \cancel{G}, \underline{S}, L$.

Joint distribution:

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

Goal: $P(J)$.

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



Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \cancel{G}, \underline{S}, L$.

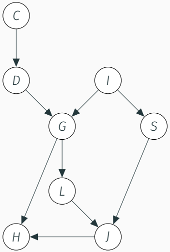
Joint distribution:

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

Goal: $P(J)$.

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

$$= \phi_J(J, L) \underbrace{\sum_S \tau_5(J, L, S)}_{\tau_6(J,L)}$$



Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \cancel{G}, \underline{S}, L$.

Joint distribution:

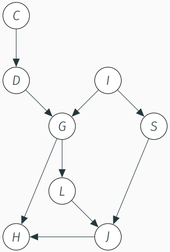
$$P(C, D, I, G, S, L, JH) = \phi_C(C)\phi_D(D, C)\phi_I(I)\phi_S(S, I) \times \phi_G(G, D, I)\phi_L(L, G)\phi_J(J, L)\phi_H(H, G, J)$$

Goal: $P(J)$.

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

$$= \phi_J(J, L)\underbrace{\sum_S \tau_5(J, L, S)}_{\tau_6(J,L)}$$

$$= \phi_J(J, L)\tau_6(J, L)$$



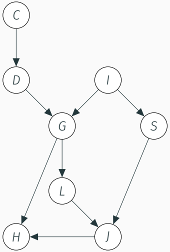
Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \cancel{G}, \cancel{S}, \underline{L}$.

Joint distribution:

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

Goal: $P(J)$.

$$\sum_L \overbrace{\phi_J(J, L)\tau_6(J, L)}^{\rho_7(J,L)}$$



Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \cancel{G}, \cancel{S}, \underline{L}$.

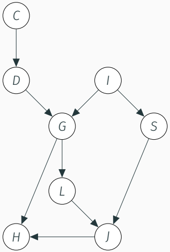
Joint distribution:

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

Goal: $P(J)$.

$$\sum_L \overbrace{\phi_J(J, L)\tau_6(J, L)}^{\rho_7(J,L)}$$

$$= \tau_7(J)$$



Eliminate: $\cancel{C}, \cancel{D}, \cancel{I}, \cancel{H}, \cancel{G}, \cancel{S}, \underline{L}$.

Joint distribution:

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

Goal: $P(J)$.

$$\sum_L \overbrace{\phi_J(J, L)\tau_6(J, L)}^{\rho_7(J,L)}$$

$$= \tau_7(J)$$

$\tau_7(J)$ is precisely $P(J)$.

VE – Overview of Steps

We have a collection of factors Φ.
(They may be reduced to account for some evidence.)

For each rv Y that we want to marginalise, we

1. gather all relevant factors—those for which Y is in the scope—removing them from Φ;
2. take the product of the relevant factors;
3. marginalise Y out of the product;
4. add the resulting factor in the collection.

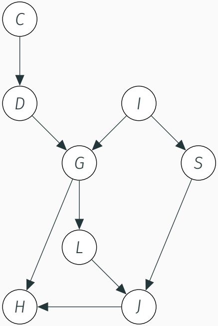
For example, marginalising G after having marginalised C, D, I, H

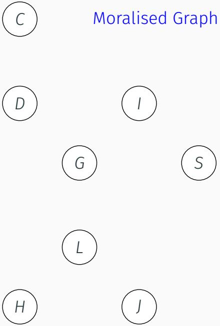
$$\sum_G \underbrace{\phi_L(L,G)\phi_J(J,L)\tau_3(G,S)\tau_4(G,J)}_{\text{(1) relevant}} = \phi_J(J,L) \underbrace{\sum_G \overbrace{\phi_L(L,G)\tau_3(G,S)\tau_4(G,J)}^{\text{(2) product}}}_{\text{(3) marginal: } \tau_5(J,L,S)} = \underbrace{\phi_J(J,L)\tau_5(J,L,S)}_{\text{(4) new collection}}$$

We repeat this until we are done marginalising.

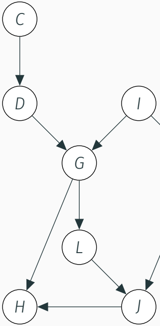
Graph View: Moralised Graph

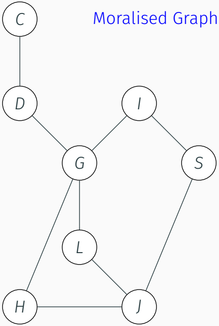
Extended Student





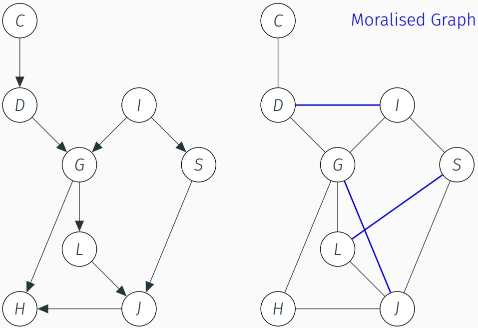
The **Moralised Graph** [definition 4.16] captures the factor structure of the BN:





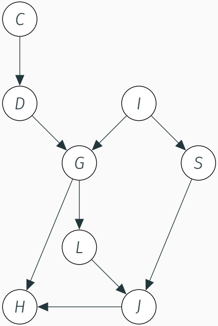
The **Moralised Graph** [definition 4.16] captures the factor structure of the BN:

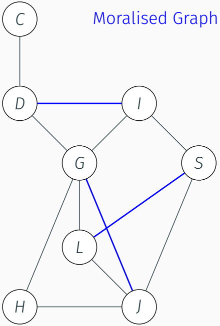
1. make edges undirected



The **Moralised Graph** [definition 4.16] captures the factor structure of the BN:

1. make edges undirected
2. *'marry'* the parents of the colliders



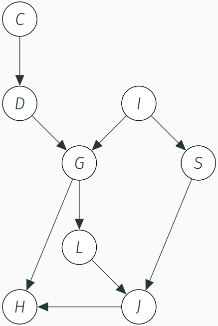


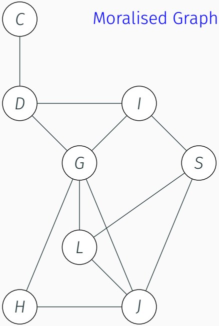
The **Moralised Graph** [definition 4.16] captures the factor structure of the BN:

1. make edges undirected
2. *'marry'* the parents of the colliders

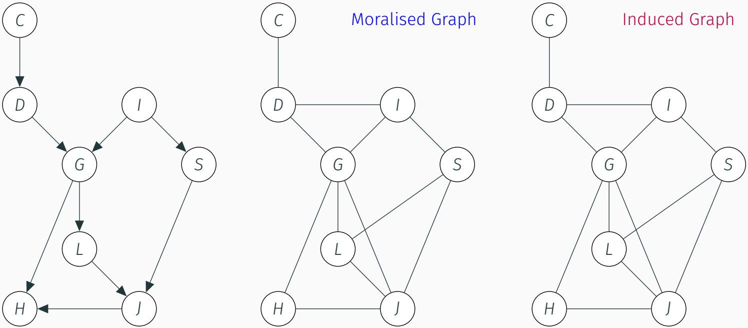
For BNs, VE starts from this graph.

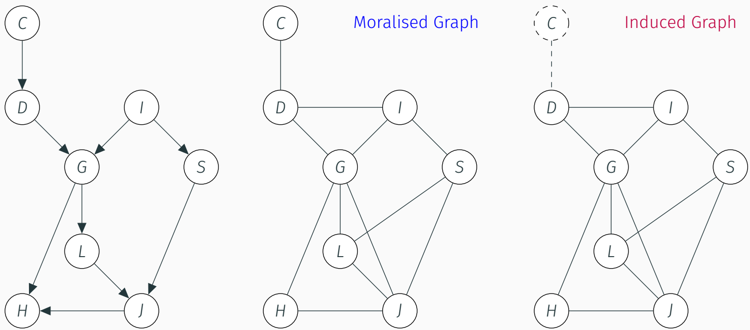
$P(J)$ for *Extended Student* [table 9.1] — Graph View



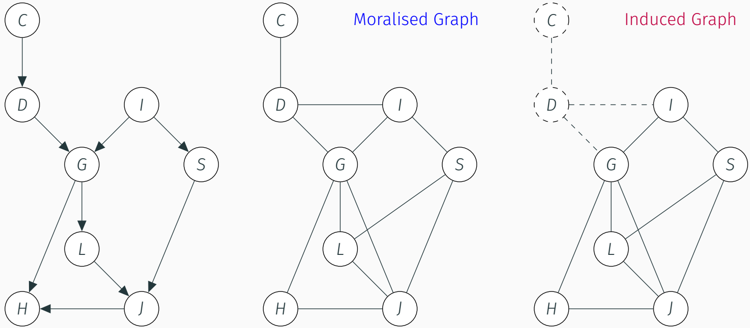


$P(J)$ for *Extended Student* [table 9.1]



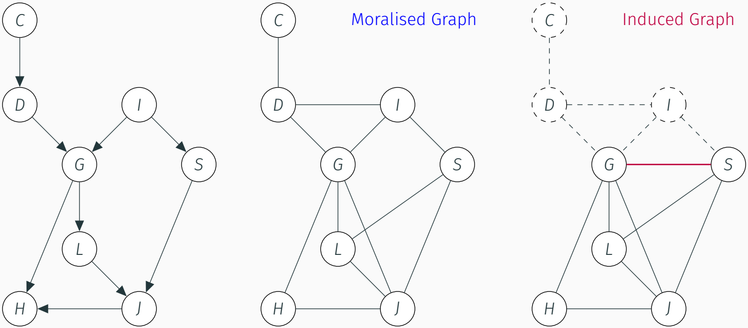


Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$

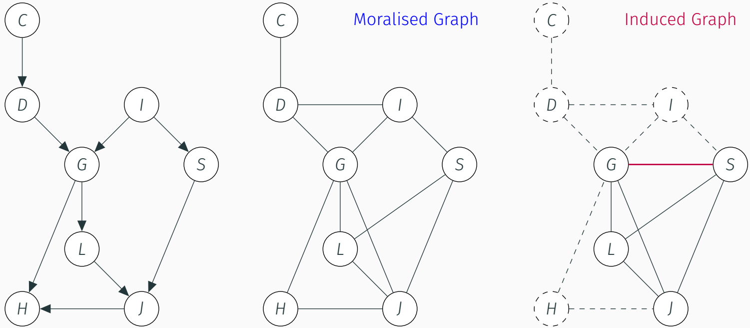


Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$
D	$\phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J, \tau_1$	τ_1, ϕ_G	$\rho_2(G, D, I)$	$\tau_2(G, I)$

$P(J)$ for *Extended Student* [table 9.1] — Graph View

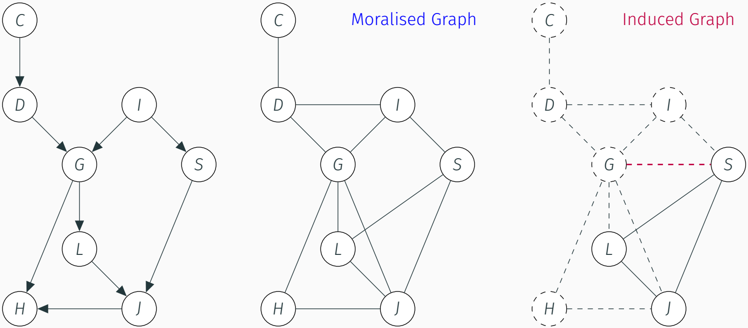


Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$
D	$\phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J, \tau_1$	τ_1, ϕ_G	$\rho_2(G, D, I)$	$\tau_2(G, I)$
I	$\phi_I, \phi_S, \phi_L, \phi_H, \phi_J, \tau_2$	ϕ_I, ϕ_S, τ_2	$\rho_3(I, G, S)$	$\tau_3(G, S)$

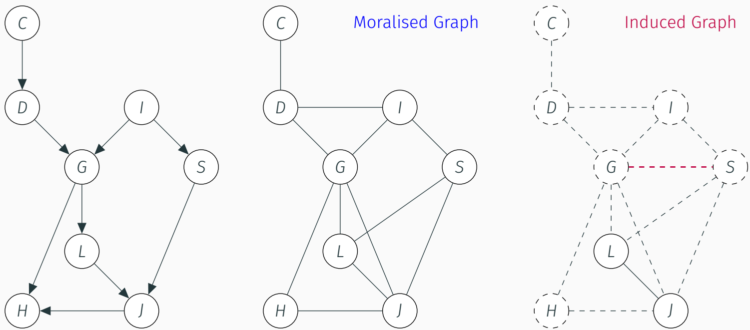


Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$
D	$\phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J, \tau_1$	τ_1, ϕ_G	$\rho_2(G, D, I)$	$\tau_2(G, I)$
I	$\phi_I, \phi_S, \phi_L, \phi_H, \phi_J, \tau_2$	ϕ_I, ϕ_S, τ_2	$\rho_3(I, G, S)$	$\tau_3(G, S)$
H	$\phi_L, \phi_H, \phi_J, \tau_3$	ϕ_H	$\phi_H(H, G, J)$	$\tau_4(G, J)$

$P(J)$ for *Extended Student* [table 9.1]

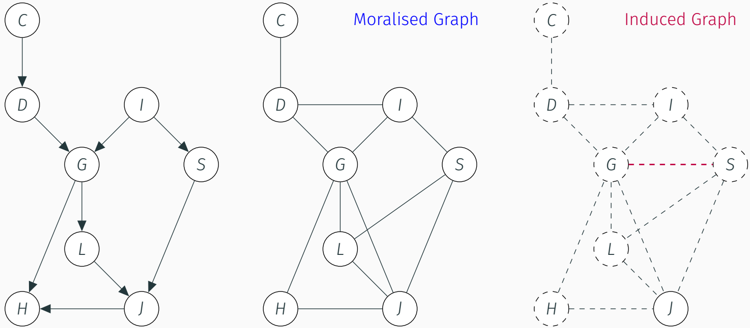


Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$
D	$\phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J, \tau_1$	τ_1, ϕ_G	$\rho_2(G, D, I)$	$\tau_2(G, I)$
I	$\phi_I, \phi_S, \phi_L, \phi_H, \phi_J, \tau_2$	ϕ_I, ϕ_S, τ_2	$\rho_3(I, G, S)$	$\tau_3(G, S)$
H	$\phi_L, \phi_H, \phi_J, \tau_3$	ϕ_H	$\phi_H(H, G, J)$	$\tau_4(G, J)$
G	$\phi_L, \phi_J, \tau_3, \tau_4$	ϕ_L, τ_3, τ_4	$\rho_5(G, J, L, S)$	$\tau_5(J, L, S)$

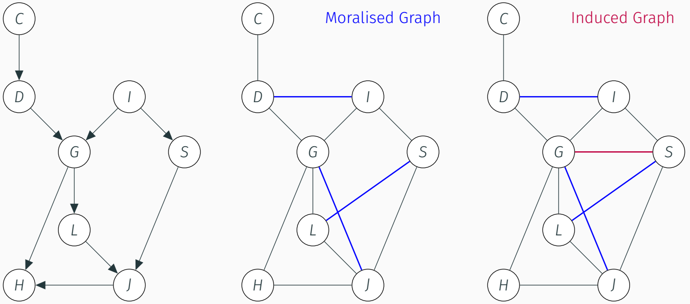


Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$
D	$\phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J, \tau_1$	τ_1, ϕ_G	$\rho_2(G, D, I)$	$\tau_2(G, I)$
I	$\phi_I, \phi_S, \phi_L, \phi_H, \phi_J, \tau_2$	ϕ_I, ϕ_S, τ_2	$\rho_3(I, G, S)$	$\tau_3(G, S)$
H	$\phi_L, \phi_H, \phi_J, \tau_3$	ϕ_H	$\phi_H(H, G, J)$	$\tau_4(G, J)$
G	$\phi_L, \phi_J, \tau_3, \tau_4$	ϕ_L, τ_3, τ_4	$\rho_5(G, J, L, S)$	$\tau_5(J, L, S)$
S	ϕ_J, τ_5	τ_5	$\tau_5(J, L, S)$	$\tau_6(J, L)$

$P(J)$ for *Extended Student* [table 9.1] Graph View



Eliminating	Available factors	Relevant	Intermediate factor	New factor
C	$\phi_C, \phi_D, \phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J$	ϕ_C, ϕ_D	$\rho_1(C, D)$	$\tau_1(D)$
D	$\phi_I, \phi_S, \phi_G, \phi_L, \phi_H, \phi_J, \tau_1$	τ_1, ϕ_G	$\rho_2(G, D, I)$	$\tau_2(G, I)$
I	$\phi_I, \phi_S, \phi_L, \phi_H, \phi_J, \tau_2$	ϕ_I, ϕ_S, τ_2	$\rho_3(I, G, S)$	$\tau_3(G, S)$
H	$\phi_L, \phi_H, \phi_J, \tau_3$	ϕ_H	$\phi_H(H, G, J)$	$\tau_4(G, J)$
G	$\phi_L, \phi_J, \tau_3, \tau_4$	ϕ_L, τ_3, τ_4	$\rho_5(G, J, L, S)$	$\tau_5(J, L, S)$
S	ϕ_J, τ_5	τ_5	$\tau_5(J, L, S)$	$\tau_6(J, L)$
L	ϕ_J, τ_6	ϕ_J, τ_6	$\rho_7(J, L)$	$\tau_7(J)$

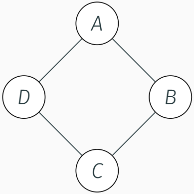


Elimination of Y as graph transformation [section 9.4.2.2]:

1. connect all neighbours of Y to one another—this captures the marginal factor that we add to the collection; depending on elimination order, this may create new edges called 'fill edges';
2. remove Y and all of its incident edges from the graph;

The **Induced Graph** [definition 9.5] is the union of all intermediate graphs.

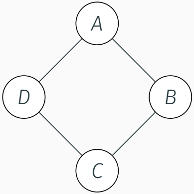
Exercise: elimination order G, I, S, L, H, C, D [solution in table 9.2].



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

$$\tilde{P}_{\Phi}(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

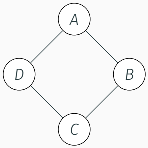


Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

$$\tilde{P}_{\Phi}(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

Example: let's express $P_{\Phi}(D)$ by eliminating A, B, C in this order



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

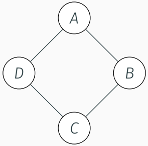
Eliminate: $\underline{A}, B, C$.

Joint unnormalised distribution:

$$\tilde{P}_{\mathbf{\Phi}}(A,B,C,D) = \phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(D,A)$$

Goal: $P_{\mathbf{\Phi}}(D)$.

$$\sum_A \phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(D,A)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

$$\tilde{P}_\Phi(A,B,C,D) = \phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(D,A)$$

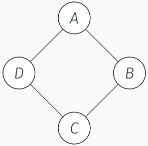
Goal: $P_\Phi(D)$.

Eliminate: $\underline{A}, B, C$.

$$\sum_A \phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(D,A)$$

$$= \phi_2(B,C)\phi_3(C,D)\underbrace{\sum_A \overbrace{\phi_1(A,B)\phi_4(D,A)}^{\rho_1(A,B,D)}}_{\tau_1(B,D)}$$

$P_{\Phi}(D)$ for *Misconception* — Eliminate A



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

$$\tilde{P}_{\Phi}(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

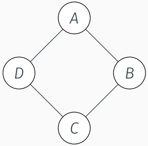
Goal: $P_{\Phi}(D)$.

Eliminate: $\underline{A}, B, C$.

$$\sum_A \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

$$= \phi_2(B, C)\phi_3(C, D) \underbrace{\sum_A \overbrace{\phi_1(A, B)\phi_4(D, A)}^{\rho_1(A,B,D)}}_{\tau_1(B,D)}$$

$$= \phi_2(B, C)\phi_3(C, D)\tau_1(B, D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

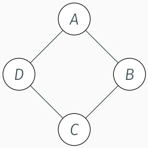
Joint unnormalised distribution:

$$\tilde{P}_\Phi(A, B, C, D) \\ = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

Goal: $P_\Phi(D)$.

Eliminate: $\cancel{A}$, $\underline{B}$, C.

$$\sum_B \phi_2(B, C)\phi_3(C, D)\tau_1(B, D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

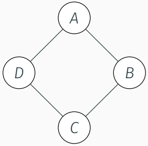
$$\tilde{P}_\Phi(A,B,C,D) = \phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(D,A)$$

Goal: $P_\Phi(D)$.

Eliminate: $\cancel{A}$, $\underline{B}$, C.

$$\sum_B \phi_2(B,C)\phi_3(C,D)\tau_1(B,D)$$

$$= \phi_3(C,D) \underbrace{\sum_B \overbrace{\phi_2(B,C)\tau_1(B,D)}^{\rho_2(B,C,D)}}_{\tau_2(C,D)}$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

$$\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

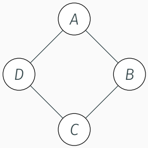
Goal: $P_\Phi(D)$.

Eliminate: $\cancel{A}$, $\underline{B}$, C.

$$\sum_B \phi_2(B, C)\phi_3(C, D)\tau_1(B, D)$$

$$= \underbrace{\phi_3(C, D) \sum_B \overbrace{\phi_2(B, C)\tau_1(B, D)}^{\rho_2(B,C,D)}}_{\tau_2(C,D)}$$

$$= \phi_3(C, D)\tau_2(C, D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

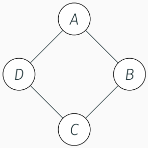
Joint unnormalised distribution:

$$\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

Goal: $P_\Phi(D)$.

Eliminate: $\cancel{A}$, $\cancel{B}$, $\underline{C}$.

$$\sum_C \phi_3(C, D)\tau_2(C, D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Eliminate: $\cancel{A}$, $\cancel{B}$, $\underline{C}$.

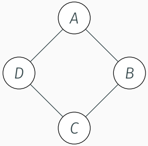
Joint unnormalised distribution:

$$\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

Goal: $P_\Phi(D)$.

$$\sum_C \phi_3(C, D)\tau_2(C, D)$$

$$= \sum_C \overbrace{\phi_3(C, D)\tau_2(C, D)}^{\rho_3(C,D)}$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

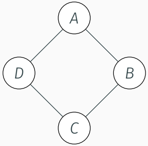
Eliminate: $\cancel{A}$, $\cancel{B}$, $\underline{C}$.

Joint unnormalised distribution:

$$\tilde{P}_\Phi(A, B, C, D) \\ = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

Goal: $P_\Phi(D)$.

$$\sum_C \phi_3(C, D)\tau_2(C, D)$$
$$= \sum_C \overbrace{\phi_3(C, D)\tau_2(C, D)}^{\rho_3(C,D)}$$
$$= \tau_3(D)$$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

Joint unnormalised distribution:

$$\tilde{P}_\Phi(A, B, C, D) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$$

Goal: $P_\Phi(D)$.

Eliminate: $\cancel{A}$, $\cancel{B}$, $\underline{C}$.

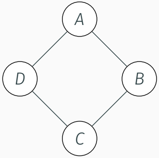
$$\sum_C \phi_3(C, D)\tau_2(C, D)$$

$$= \sum_C \overbrace{\phi_3(C, D)\tau_2(C, D)}^{\rho_3(C,D)}$$

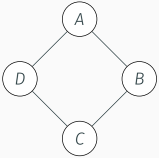
$$= \tau_3(D)$$

$\tau_3(D)$ is $\tilde{P}_\Phi(D)$.

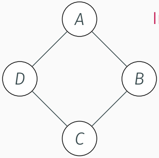
To obtain $P_\Phi(D)$ we normalise $\tilde{P}(D)$ using the normaliser $\sum_{d \in \mathrm{Val}(D)} \tau_3(D = d)$.



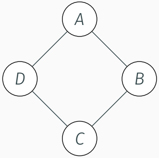
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A



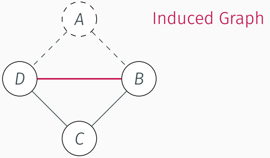
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A



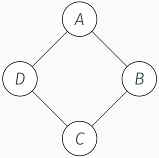
Induced Graph



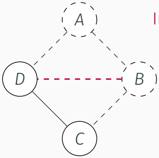
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A



Eliminating	Available factors	Relevant	Intermediate factor	New factor
A	$\phi_1, \phi_2, \phi_3, \phi_4$	ϕ_1, ϕ_4	$\rho_1(A, B, D)$	$\tau_1(B, D)$

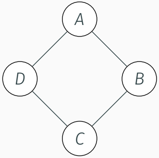


Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

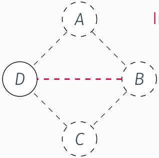


Eliminating	Available factors	Relevant	Intermediate factor	New factor
A	$\phi_1, \phi_2, \phi_3, \phi_4$	ϕ_1, ϕ_4	$\rho_1(A, B, D)$	$\tau_1(B, D)$
B	ϕ_2, ϕ_3, τ_1	ϕ_2, τ_1	$\rho_2(B, C, D)$	$\tau_2(C, D)$

$P_\Phi(D)$ for *Misconception* — Graph View

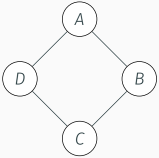


Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A

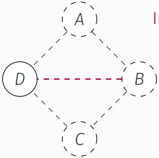


Induced Graph

Eliminating	Available factors	Relevant	Intermediate factor	New factor
A	$\phi_1, \phi_2, \phi_3, \phi_4$	ϕ_1, ϕ_4	$\rho_1(A, B, D)$	$\tau_1(B, D)$
B	ϕ_2, ϕ_3, τ_1	ϕ_2, τ_1	$\rho_2(B, C, D)$	$\tau_2(C, D)$
C	ϕ_3, τ_2	ϕ_3, τ_2	$\rho_3(C, D)$	$\tau_3(D)$



Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A



Eliminating	Available factors	Relevant	Intermediate factor	New factor
A	$\phi_1, \phi_2, \phi_3, \phi_4$	ϕ_1, ϕ_4	$\rho_1(A, B, D)$	$\tau_1(B, D)$
B	ϕ_2, ϕ_3, τ_1	ϕ_2, τ_1	$\rho_2(B, C, D)$	$\tau_2(C, D)$
C	ϕ_3, τ_2	ϕ_3, τ_2	$\rho_3(C, D)$	$\tau_3(D)$

There's no need for 'moralisation' because MNs do not have v-structures.

Elimination of Y as graph transformation [section 9.4.2.2]:

1. connect all neighbours of Y to one another;
2. remove Y and all of its incident edges from the graph;

The **Induced Graph** [definition 9.5] is the union of all intermediate graphs.

Exercise: VE in order B, C, A.

For a BN with nodes $\boldsymbol{X}$, start from the moralised graph with factors Φ.
For an MN with nodes $\boldsymbol{X}$, start from its factors Φ.

If there's evidence $\boldsymbol{E} = \boldsymbol{e}$, reduce all factors in Φ. [Section 9.3.2]

For a query $\boldsymbol{Q}$, disjoint with $\boldsymbol{E}$, we need to marginalise $\boldsymbol{M} = \boldsymbol{X} \setminus (\boldsymbol{Q} \cup \boldsymbol{E})$.

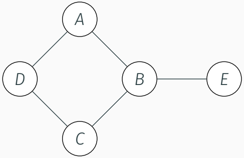
In a given order α, eliminate the variables in $\boldsymbol{M}$ one by one.

For each rv $Y \in \boldsymbol{M}$ that we want to marginalise, we

1. gather all relevant factors—those for which Y is in the scope—removing them from Φ;
2. take the product of the relevant factors;
3. marginalise Y out of the product;
4. add the resulting factor in the collection.

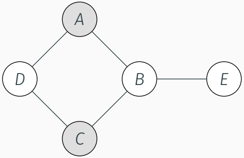
At the end, we take the product of whatever factors remain and normalise to obtain a distribution.

VE with Evidence: $P_\Phi(D|A = a^1, C = c^1)$ Graph View



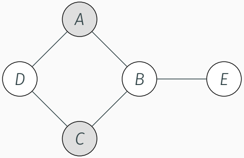
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A
ϕ_5	B, E

VE with Evidence: $P_\Phi(D|A = a^1, C = c^1)$ — Graph View

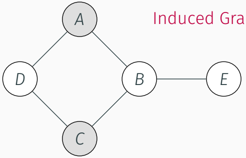


Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E

VE with Evidence: $P_\Phi(D|A = a^1, C = c^1)$ Graph View

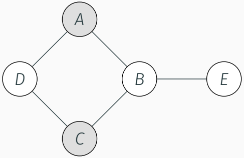


Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E

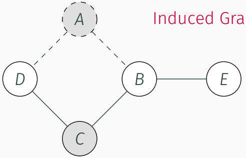


Induced Graph

VE with Evidence: $P_\Phi(D|A = a^1, C = c^1)$ — Graph View

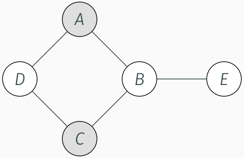


Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E

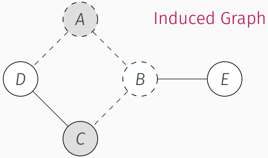


Induced Graph

Elim.	Available factors	Relevant	Intermediate	New factor
A	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\emptyset$	-	-

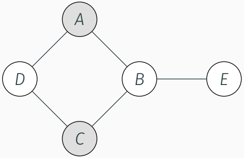


Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E

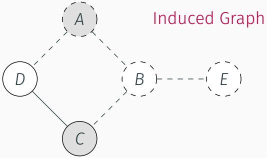


Elim.	Available factors	Relevant	Intermediate	New factor
A	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\emptyset$	-	-
B	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\phi_1[a^1], \phi_2[c^1], \phi_5$	$\rho_1(B, E)$	$\tau_1(E)$

VE with Evidence: $P_\Phi(D|A = a^1, C = c^1)$

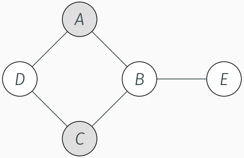


Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E

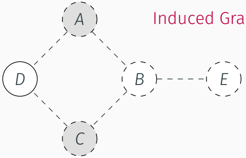


Elim.	Available factors	Relevant	Intermediate	New factor
A	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\emptyset$	-	-
B	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\phi_1[a^1], \phi_2[c^1], \phi_5$	$\rho_1(B, E)$	$\tau_1(E)$
E	$\phi_3[c^1], \phi_4[a^1], \tau_1$	τ_1	$\tau_1(E)$	γ_1

VE with Evidence: $P_\Phi(D|A=a^1, C=c^1)$ — Graph View



Φ	Scope
$\phi_1[A=a^1]$	B
$\phi_2[C=c^1]$	B
$\phi_3[C=c^1]$	D
$\phi_4[A=a^1]$	D
ϕ_5	B, E

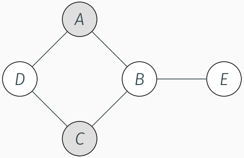


Induced Graph

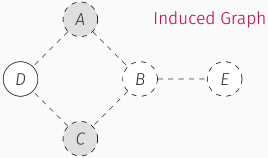
Elim.	Available factors	Relevant	Intermediate	New factor
A	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\emptyset$	-	-
B	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\phi_1[a^1], \phi_2[c^1], \phi_5$	$\rho_1(B, E)$	$\tau_1(E)$
E	$\phi_3[c^1], \phi_4[a^1], \tau_1$	τ_1	$\tau_1(E)$	γ_1
C	$\phi_3[c^1], \phi_4[a^1]\gamma_1$	$\emptyset$	-	-

$$\tilde{P}_\Phi(D, A=a^1, C=c^1) = \phi_3[c^1](D)\phi_4[a^1](D)\gamma_1$$

VE with Evidence: $P_\Phi(D|A = a^1, C = c^1)$ — Graph View



Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E



Elim.	Available factors	Relevant	Intermediate	New factor
A	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\emptyset$	-	-
B	$\phi_1[a^1], \phi_2[c^1], \phi_3[c^1], \phi_4[a^1], \phi_5$	$\phi_1[a^1], \phi_2[c^1], \phi_5$	$\rho_1(B, E)$	$\tau_1(E)$
E	$\phi_3[c^1], \phi_4[a^1], \tau_1$	τ_1	$\tau_1(E)$	γ_1
C	$\phi_3[c^1], \phi_4[a^1]\gamma_1$	$\emptyset$	-	-

$$\tilde{P}_\Phi(D, A = a^1, C = c^1) = \phi_3[c^1](D)\phi_4[a^1](D)\gamma_1$$

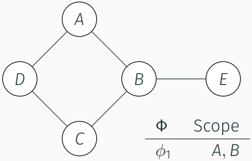
To obtain $P_\Phi(D|A = a^1, C = c^1)$ we divide $\tilde{P}_\Phi(D, A = a^1, C = c^1)$ by the normaliser $\sum_{d \in \mathrm{Val}(D)} \phi_3[c^1](D = d)\phi_4[a^1](D = d)\gamma_1$.

Graphical Separation

We can use graphical separation to spare some unnecessary computation in VE.

Key: if an rv Y to be eliminate is separate from Q given E, then we can eliminate Y without multiplying the factors that involve Y.

Example:



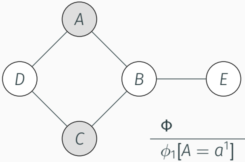
Φ	Scope
ϕ_1	A, B
ϕ_2	B, C
ϕ_3	C, D
ϕ_4	D, A
ϕ_5	B, E

Graphical Separation

We can use graphical separation to spare some unnecessary computation in VE.

Key: if an rv Y to be eliminate is separate from Q given E, then we can eliminate Y without multiplying the factors that involve Y.

Example:

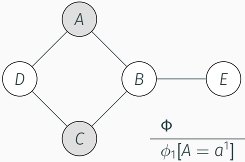


Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E

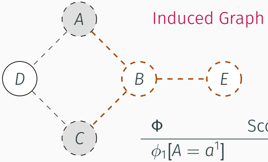
We can use graphical separation to spare some unnecessary computation in VE.

Key: if an rv Y to be eliminate is separate from Q given E, then we can eliminate Y without multiplying the factors that involve Y.

Example:



Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E



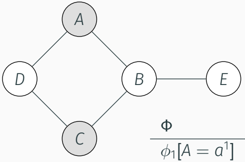
Φ	Scope	Available
$\phi_1[A = a^1]$	B	✗
$\phi_2[C = c^1]$	B	✗
$\phi_3[C = c^1]$	D	✓
$\phi_4[A = a^1]$	D	✓
ϕ_5	B, E	✗

Graphical Separation

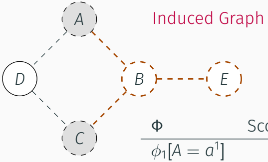
We can use graphical separation to spare some unnecessary computation in VE.

Key: if an rv Y to be eliminate is separate from Q given E, then we can eliminate Y without multiplying the factors that involve Y.

Example:



Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E



Φ	Scope	Available
$\phi_1[A = a^1]$	B	✗
$\phi_2[C = c^1]$	B	✗
$\phi_3[C = c^1]$	D	✓
$\phi_4[A = a^1]$	D	✓
ϕ_5	B, E	✗

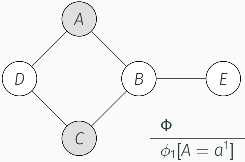
$$\tilde{P}_{\Phi}(D, A = a^1, C = c^1) = \phi_3[c^1](D)\phi_4[a^1](D)$$

Graphical Separation

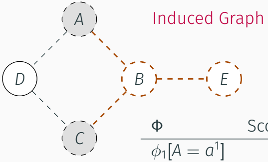
We can use graphical separation to spare some unnecessary computation in VE.

Key: if an rv Y to be eliminate is **separate** from Q given E, then we can eliminate Y without multiplying the factors that involve Y.

Example:



Φ	Scope
$\phi_1[A = a^1]$	B
$\phi_2[C = c^1]$	B
$\phi_3[C = c^1]$	D
$\phi_4[A = a^1]$	D
ϕ_5	B, E



Φ	Scope	Available
$\phi_1[A = a^1]$	B	✗
$\phi_2[C = c^1]$	B	✗
$\phi_3[C = c^1]$	D	✓
$\phi_4[A = a^1]$	D	✓
ϕ_5	B, E	✗

$\tilde{P}_{\Phi}(D, A = a^1, C = c^1) = \phi_3[c^1](D)\phi_4[a^1](D)$. This induces the same $P_{\Phi}(D|A = a^1, C = c^1)$ as before because γ_1 is constant.

Each factor in Φ is used *once*.

Each factor in Φ is used *once*.

The **number of new factors is linear** in the size of M.

Each factor in Φ is used *once*.

The **number of new factors is linear** in the size of M.

Every factor formed in VE (incl. intermediate factors) corresponds to a *clique in the induced graph*.

Each factor in Φ is used *once*.

The **number of new factors is linear** in the size of M.

Every factor formed in VE (incl. intermediate factors) corresponds to a *clique in the induced graph*.

The **largest factor** formed has the size of the largest *maximal* clique in the induced graph. A clique is maximal if there's no way to expand it into a larger subgraph that's still a clique.

Each factor in Φ is used *once*.

The **number of new factors is linear** in the size of M.

Every factor formed in VE (incl. intermediate factors) corresponds to a *clique in the induced graph*.

The **largest factor** formed has the size of the largest *maximal* clique in the induced graph. A clique is maximal if there's no way to expand it into a larger subgraph that's still a clique.

If the largest factor has N rvs in scope, each with cardinality K, then the **complexity of VE is proportional to** K^N.

Each factor in Φ is used *once*.

The **number of new factors is linear** in the size of M.

Every factor formed in VE (incl. intermediate factors) corresponds to a *clique in the induced graph*.

The **largest factor** formed has the size of the largest *maximal* clique in the induced graph. A clique is maximal if there's no way to expand it into a larger subgraph that's still a clique.

If the largest factor has N rvs in scope, each with cardinality K, then the **complexity of VE is proportional to** K^N.

The worst case of VE is intractable, much like naive sum-product, but because most models have *sparse graph structure*, **VE is typically much better than naive sum-product**.

Each factor in Φ is used *once*.

The **number of new factors is linear** in the size of M.

Every factor formed in VE (incl. intermediate factors) corresponds to a *clique in the induced graph*.

The **largest factor** formed has the size of the largest *maximal* clique in the induced graph. A clique is maximal if there's no way to expand it into a larger subgraph that's still a clique.

If the largest factor has N rvs in scope, each with cardinality K, then the **complexity of VE is proportional to** K^N.

The worst case of VE is intractable, much like naive sum-product, but because most models have *sparse graph structure*, **VE is typically much better than naive sum-product**.

The *largest factor in VE depends on the order of elimination*. But, there's no polynomial-time algorithm to find an optimal ordering.

Summary

- Solving marginals is the key to complex probability queries (with or without evidence).
- Naive sum-product builds the largest possible factor—the (unnormalised) joint distribution—then factor marginalisation is safe to use. This algorithm is intractable for all but the smallest PGMs.
- VE marginalises ('eliminates') one rv at a time in a given order. By taking products involving only the necessary factors, VE achieves efficient inference (so long as the graph is sparse).

What Next?

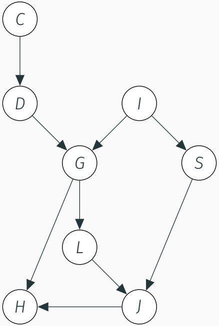
LC4: sum-product VE in code.

HC4b: elimination orderings.

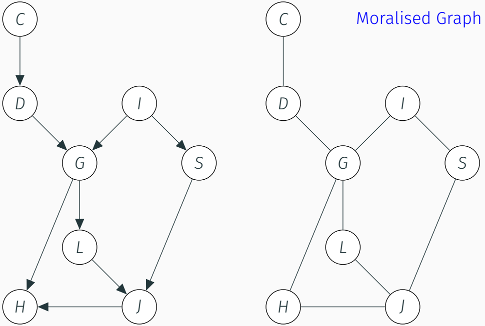
WC4: exercises.

Elimination Ordering

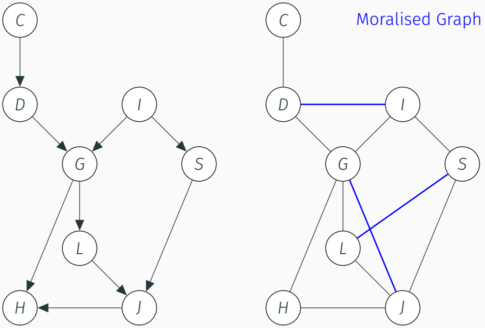
Let's run VE to express $P(J)$ and let's start with eliminating G



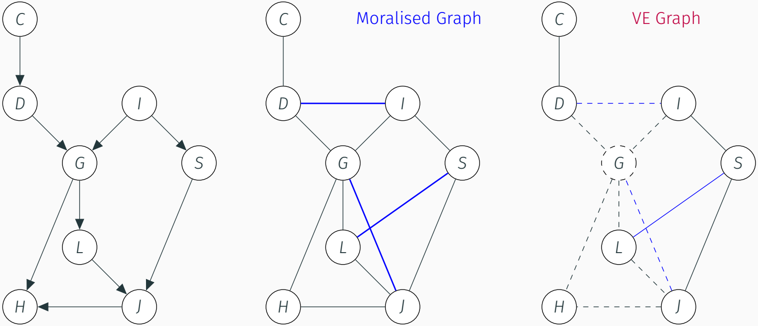
Let's run VE to express $P(J)$ and let's start with eliminating G



Let's run VE to express $P(J)$ and let's start with eliminating G

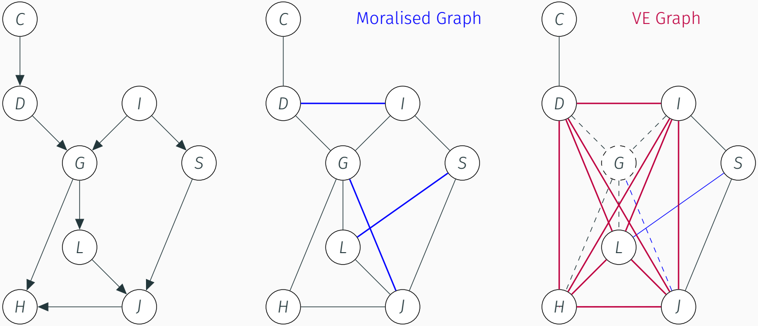


Let's run VE to express $P(J)$ and let's start with eliminating G



All cliques involving G are dashed. Some or all of them will have factors in Φ, we should identify those factors and take their product.

Let's run VE to express $P(J)$ and let's start with eliminating G



All cliques involving G are dashed. Some or all of them will have factors in Φ, we should identify those factors and take their product. This creates a **large intermediate factor** with scope D, G, H, I, J, L, which has cardinality 3×2^5. In example 9.1, the largest intermediate factor had scope G, J, L, S, which has cardinality 3×2^3

Optimal Ordering

An optimal ordering is one for which the largest intermediate factor is as small as possible.

Unfortunately, finding such an order is intractable. [Section 9.4.3]

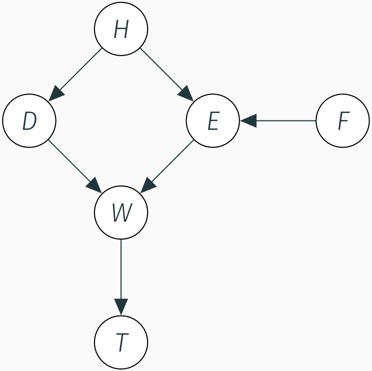
We resort to some useful heuristics: [Section 9.4.3.2]

- Min-neighbours
- Min-weight
- Min-fill
- Weighted-min-fill

These work surprisingly well in practice (min-fill and weighted-min-fill tend to work better on more problems).

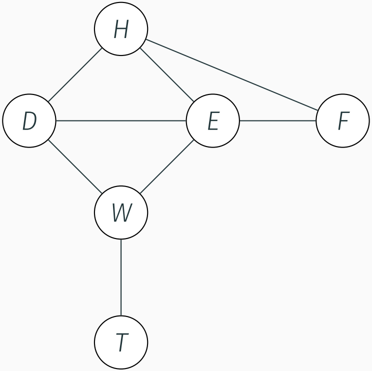
Min-Neighbours

The cost of a node is the number of neighbours it has.



Min-Neighbours

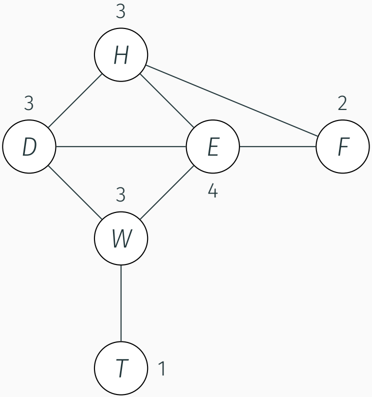
The cost of a node is the number of neighbours it has.



Moralise the BN and count neighbours.

Min-Neighbours

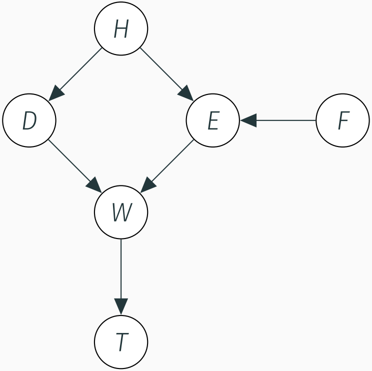
The cost of a node is the number of neighbours it has.



Moralise the BN and count neighbours.
Break ties lexicographically: T, F, D, H, W, E.

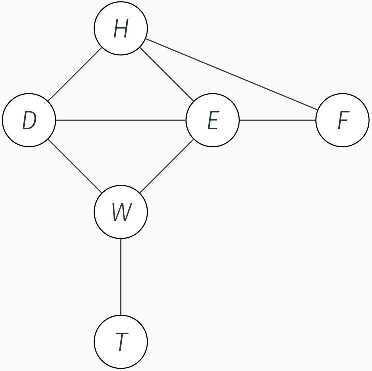
Min-Weight

The cost of a node is the product of *weights* (domain cardinality) of its neighbours.



Min-Weight

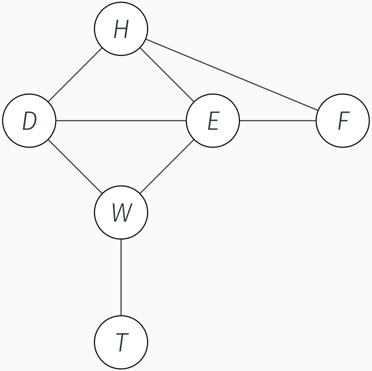
The cost of a node is the product of *weights* (domain cardinality) of its neighbours.



Moralise the BN and quantify weight.

Min-Weight

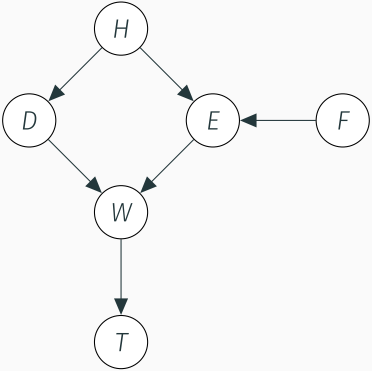
The cost of a node is the product of *weights* (domain cardinality) of its neighbours.



Moralise the BN and quantify weight.
Break ties lexicographically:

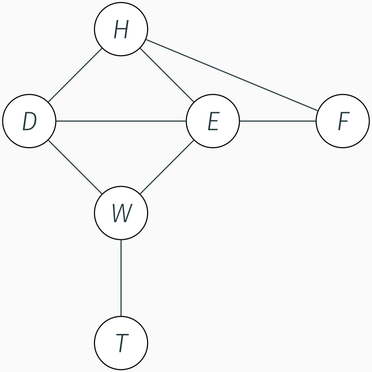
Min-Fill

The cost of a node is the number of edges that need to be *added* to the graph due to its elimination.



Min-Fill

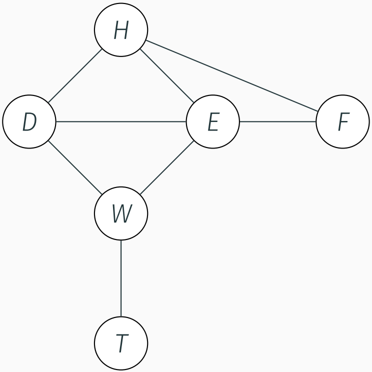
The cost of a node is the number of edges that need to be *added* to the graph due to its elimination.



Moralise the BN and count fill edges.

Min-Fill

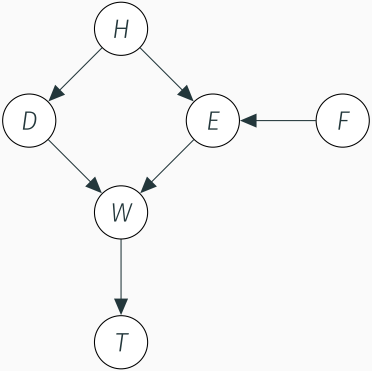
The cost of a node is the number of edges that need to be *added* to the graph due to its elimination.



Moralise the BN and count fill edges.
Break ties lexicographically:

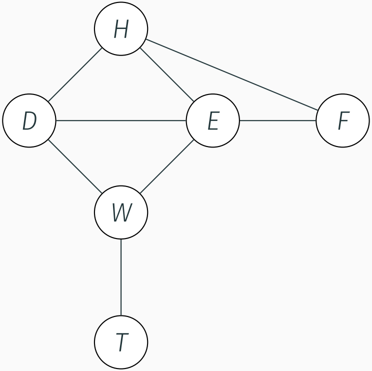
Weighted-Min-Fill

The cost of a node is the sum of weights of the edges that need to be *added* to the graph due to its elimination, where weight of an edge is the product of weights of its vertices.



Weighted-Min-Fill

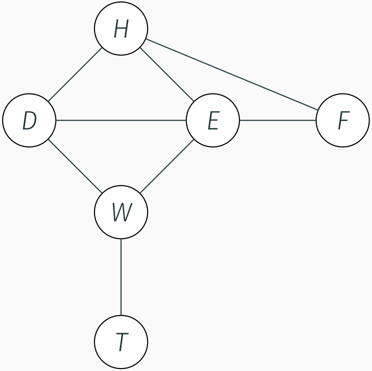
The cost of a node is the sum of weights of the edges that need to be *added* to the graph due to its elimination, where weight of an edge is the product of weights of its vertices.



Moralise the BN and quantify total weight of fill edges.

Weighted-Min-Fill

The cost of a node is the sum of weights of the edges that need to be *added* to the graph due to its elimination, where weight of an edge is the product of weights of its vertices.



Moralise the BN and quantify total weight of fill edges.
Break ties lexicographically:

Induced Graph and Search for Elimination Ordering

Suppose we have 4 heuristics (such as those just presented), we can simulate their corresponding induced graphs in VE (that is, we draw the induced graph without actually taking any factor product).

Then we can find out which heuristic leads to the most efficient VE run: the one whose induced graph requires the least prohibitive factor products.

Exercises

Summary

- The most efficient VE run is the order that creates the smallest maximal clique.
- Optimal orderings are intractable, but we a number of heuristics work well in practice.
- We can even simulate the effect of different heuristics by working out their induced graphs.

What Next?

WC4: exercises.

Next module: approximate inference.

References

[1] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.